Symmetrically Censored GMM Estimation
for Tobit Models with Endogenous Regressors∗

Haizheng Li

School of Economics
Georgia Institute of Technology
Atlanta, GA 30332-0615
Phone: (404) 894-3542
Fax: (404) 894-1890
haizheng.li@econ.gatech.edu

∗I would like to thank Jeffrey Zax, Jushan Bai, Shinichi Sakata, Whitney Newey, Donald Waldman, Robert McNown, Nicholas Flores, B.J. Lee, Songnian Chen, Buhong Zheng, Thomas Dee, Lynn Yang, and especially two anonymous referees for their comments. Seminar participants at the 1996 North American Summer Meetings of the Econometric Society, Ohio Sate University, University of Colorado-Boulder, McGill University, Georgia Institute of Technology, and University of Missouri-Columbia have been generous with comments.
Symmetrically Censored GMM Estimation
for Tobit Models with Endogenous Regressors

Abstract

In this study, a new semi-parametric estimation method for censored models, symmetrically censored generalized method of moments (SCGMM) estimator, is proposed. The SCGMM estimator can be applied to censored models when some regressors are correlated with the regression error. Moreover, it does not impose the independence assumption for data, and thus can be applied to time-series and panel data models with dependent and heterogeneous observations. Furthermore, the SCGMM does not require any non-parametric techniques for estimation, and is relatively simple to implement. The SCGMM estimator is also robust to a fairly wide variety of non-normal and non-identical but symmetric disturbance distributions. This estimator adopts a new approach to identifying the true parameter by restricting the parameter space in the estimation. The SCGMM estimator is shown to be consistent and asymptotically normal for near epoch dependent functions of mixing processes. Thus, it can be applied to models with mixing processes, and models in which the disturbance term is an ARMA process of finite order (with roots outside the unit circle), or an infinite non-stationary MA process. A heteroskedasticity and autocorrelation consistent estimator of the covariance matrices is also provided. Simulation results show that the SCGMM estimator performs very well.

Keywords: Semi-parametric, censored model, GMM, near epoch dependence, endogeneity.
I. Introduction

Regression models with censored dependent variables are common in empirical work. It is known that the commonly used Tobit maximum likelihood estimation procedure for censored models is sensitive to the error term distribution. If the underlying distribution is not both normal and homoskedastic, the Tobit procedure will generally be inconsistent (Arabmazar and Schmidt 1982, Goldberger 1983). Several semi-parametric or non-parametric estimators have been proposed to relax these assumptions, for example, Powell (1984), Horowitz (1986), Powell (1986), and Honoré and Powell (1994). These estimators are robust to heteroskedasticity and non-normality of the disturbance distribution.

However, these semi-parametric estimators require that all regressors in a censored model be uncorrelated with (or independent of) the regression error. If some explanatory variables are correlated with the regression error, these estimators are generally inconsistent. This restriction rules out endogeneity and measurement error in any regressor. In addition, they also assume independence for the data in the model. For many censored models, such as labor supply functions and some demand functions (e.g., for tickets, cigarettes, etc.), endogenous explanatory variables, for example wages and prices, are common. Furthermore, for censored models with time-series or panel data, observations are unlikely to be independent.

Therefore, it is desirable to have a semi-parametric estimator that can be applied to models with endogenous regressors and dependent observations. Lewbel (1998) proposes a non-parametric estimator for the latent variable model with endogenous regressors. Hong and Tamer (2003) extended the “censored LAD” estimator proposed in Power (1984) to censored models with endogenous regressors. These estimators require non-parametric estimations of distribution functions. Additionally, they also maintain the independence assumption for the data.

In this study, a new semi-parametric estimation method, symmetrically censored generalized method of moments (SCGMM) estimator, is proposed. The SCGMM estimator can be applied to censored models when some regressors are correlated with the regression error.
Moreover, it does not impose the independence assumption for data, and thus can be applied to
time-series and panel data models with dependent and heterogeneous observations. Furthermore,
the SCGMM does not require any non-parametric techniques in the estimation, and is relatively
simple to implement.\footnote{Non-parametric estimation generally requires choosing smoothing parameters, and no general satisfactory
solution to this problem exists.} As with other semi-parametric estimators, the SCGMM is also robust to a
fairly wide variety of non-normal and non-identical but symmetric disturbance distributions.

The SCGMM method generalizes the symmetrical trimming procedure in the
symmetrically censored least squares (SCLS) estimator proposed by Powell (1986) for censored
models with only exogenous regressors. This estimation method can be directly applied to
models with linear or non-linear endogenous explanatory variables. It does not require
constructing reduced form equations. The SCGMM estimator is shown to be consistent and
asymptotically normal under very general conditions. Furthermore, a consistent estimator of the
asymptotic covariance matrices is also established in the presence of heteroskedasticity and
autocorrelation of unknown form.

The SCGMM estimation makes use of instruments in the estimation. In order to
accommodate over-identifying instruments, it adopts a generalized method of moments (GMM)
framework. However, the standard asymptotic theory of GMM is not applicable in this case
because of the non-smoothness of the SCGMM criterion function. Nevertheless, under suitable
regularity conditions outlined in the following sections, the consistency and asymptotic normality
of the SCGMM estimator can be established.

In order to ensure unique identifiability, the SCGMM estimator adopts a new approach to
identification: instead of constructing a special criterion function to identify the true parameter,
the SCGMM estimator is defined in a constrained parameter space, in which the true parameter is
the unique solution to the moment conditions. Thus, the SCGMM estimator is a constrained
optimization estimator, and mathematical programming methods are needed for estimation.
Furthermore, in order to fit a general dependence structure, the SCGMM estimator extends Hansen's GMM estimator (1982) from stationary and ergodic sequences to near epoch dependent (NED) processes. Hence it can be applied to models with m-dependent or mixing processes, and models in which the disturbance term is an ARMA process of finite order (with roots outside the unit circle), or an infinite non-stationary MA process (under appropriate mild conditions on the MA weights). Since it does not require a stationarity assumption for dependent processes, the SCGMM estimation allows for both conditional and unconditional heteroskedasticity.

The simulation results show that the SCGMM estimator performs very well, much better than the SCLS in the presence of endogenous regressors. Although the SCGMM estimation requires a stronger assumption on the instruments, such instruments are not uncommon in empirical applications. Moreover, the SCGMM estimator is robust to a variety of censored models, especially when some explanatory variables are correlated with the disturbance term. In addition, it includes the SCLS estimation as a special case.

The rest of the paper is organized as follows. In next section, the SCGMM estimation procedure is introduced. Sections III and IV provide regularity conditions for consistency and asymptotic normality. Consistent estimation of the asymptotic covariance matrices is discussed in Section V. In section VI, a small scale simulation is conducted, and section VII concludes. Technical proofs are given in the appendix.

II. The SCGMM Estimator

Consider a “true” underlying linear regression model

\[
(2.1) \quad y_t^* = x_t' \beta_0 + u_t^* \quad (t = 1,2, ..., n),
\]

where \( t \) indexes observations, \( y_t^* \) is a dependent variable, \( x_t \) is a \( k \) vector of exogenous and endogenous regressors, \( \beta \) is a \( k \) vector of unknown constant coefficients, and \( u_t^* \) is an
unobservable error. For a censored regression, only $x_t$ and $y_t = \max\{0, y_t^*\}$ are observable. \footnote{The censoring point here is assumed to be zero, though this is only a convenient normalization.}

Thus the error term is also censored as $u_t = \max\{u_t^*, -x_t'\beta_0\}$. In this case, even if all elements of $x_t$ are exogenous, the conditional expectation of the censored error $u_t$ is nonzero, i.e., $E(u_t \mid x_t) \neq 0$.

In this case, OLS estimation is inconsistent. A common alternative is the Tobit procedure, where it is necessary to specify completely the distribution of $u_t^*$ in order to adopt the likelihood-based approach. If the distribution of $u_t^*$ is unknown, or if it is heteroskedastic with unknown form, the Tobit procedure will be inconsistent.

If all elements in $x_t$ are exogenous and $u_t^*$, conditional on $x_t$, is distributed symmetrically around zero, the SCLS proposed by Powell (1986) can be applied when the distribution of $u_t^*$ is unknown and is heteroskedastic of unknown form. More specifically, in the SCLS estimation, $u_t$ is censored from the upper tail by $x_t'\beta_0$, and thus the symmetry of the distribution is restored. As a result, the new censored error becomes $u_t^* = \min\{u_t, x_t'\beta_0\}$. As the values of $y_t^*$ below zero have been censored to zero, this procedure censors the values of $y_t^*$ above $2x_t'\beta_0$ to $2x_t'\beta_0$, i.e., $y_t^* = \min\{y_t, 2x_t'\beta_0\}$, to eliminate the asymmetry of the distribution. Thus it results in the following desirable moment condition:

\begin{equation}
(2.2) \quad E(1(x_t'\beta_0>0) \cdot u_t^* \mid x_t) = 0.
\end{equation}

In this case, the OLS estimation can be applied without any further assumptions.

However, if some elements of $x_t$ are correlated with the regression error, the symmetrically censored error $u_t^*$ will not satisfy the moment conditions (2.2). Since the trimming procedure depends on $x_t$, hence the trimming itself is also endogenous. Therefore, the SCLS estimator cannot be applied. Moreover, in this case, the usual instrumental variable estimation cannot be simply used either. More specifically, suppose there exists a vector of instruments $w_t$ and $E(u_t^* \mid w_t)=0$, and that conditional on $w_t$, $u_t^*$ is distributed symmetrically around zero. The expectation of the symmetrically censored error $u_t^*$ on $w_t$, i.e., $E(1(x_t'\beta_0>0) \cdot u_t^* \mid w_t)$, will depend
on the joint distribution of $x_t$, $u_t^*$, and $w_t$, and will generally be non-zero. Therefore, the usual
moment condition needed for IV estimation, \( E(1(x_t'\beta_0>0) \cdot u_t^* \mid w_t) = 0 \), is not satisfied.\(^3\)

Nevertheless, suppose that $x_t' = (x_{1t}', x_{2t}')$, and $x_{1t}$ is $k_1$ vector of endogenous variables, $x_{2t}$
is $k_2$ vector of exogenous variables, and $\beta_0' = (\beta_{01}', \beta_{02}')$ Then model (2.1) can be written as
\[
y_t^* = x_{1t}'\beta_{01} + x_{2t}'\beta_{02} + u_t^* \quad (t = 1, 2, ..., n).
\]
Assume that there exists a $l$ vector of instruments $w_t$ that is correlated with $x_t$ but uncorrelated
with $u_t^*$, and $l \geq k$, $w_t = (w_{1t}', w_{2t}', x_{2t}')$, where $w_{1t}$ is a $k_1$ vector of instruments and $w_{2t}$ is a $(l-k_1-k_2)$
vector of instruments. In this case, \( E(u_t^* \mid w_t) = 0 \). Furthermore, for instrument $w_{1t}$, assume the
following condition holds,
\[
(2.3) \quad x_{1t}'\beta_{01} \geq w_{1t}'\beta_{01}, \quad \text{for } t=1, 2, ..., n.
\]
Define that instrument $z$ include all exogenous regressors and the subset of instruments that
satisfies the condition (2.3), i.e., $z=(w_{1t}', x_{2t}')$. Clearly, equation (2.3) implies:
\[
(2.4) \quad x_t'\beta_0 \geq z'\beta_0 \quad \text{for } t=1, 2, ..., n.
\]
Given the existence of such instruments, the trimming procedure can be performed by $z$
instead of $x$. Therefore, a new error can be obtained by censoring $u_t^*$ below $-z_t'\beta_0$ to $-z_t'\beta_0$ and
symmetrically censor $u_t^*$ above $z_t'\beta_0$ to $z_t'\beta_0$, i.e.,
\[
(2.5) \quad u_t^c = \min \{ \max \{ u_t^*, -z_t'\beta_0 \} , z_t'\beta_0 \}.
\]
The condition (2.4) can be viewed as a requirement that $-z_t'\beta_0$ falls into the region of $u_t^*$ that is not
censored by $-x_t'\beta_0$.

\(^3\) Newey (1985b) extends the SCLS to models with endogenous regressors and proposes a Two-Stage
Instrumental Variable (2SIV) estimation. In the first stage, the OLS estimation is applied to produce the
predicted value for the endogenous regressors, and in the second stage, the SCLS procedure is applied with
the endogenous regressors replaced by their predicted values. The 2SIV estimator is based on an early
version of the SCLS estimator and the criterion function is incorrect. Thus the asymptotic theory of the
2SIV estimator needs to be re-established. Moreover, in the 2SIV estimation, it essentially requires that
\( E[1(x_t'\beta_0>0) \cdot \min \{ u_t, x_t'\beta_0 \} \mid w_t]=0 \). Since the predicted value of $x_t$ depends on both $w_t$ and $x_t$, this
moment condition may also depend on joint distribution of $x_t$, $w_t$, and $u_t$. Therefore, it seems that more
justifications are needed for the 2SIV estimator.
If $u_t^*$, conditional on $w_t$, is distributed symmetrically around zero, the following moment condition can be obtained to identify the true parameters:

$$\mathbb{E}(1(z_t'\beta_0 > 0) u_t^* \mid w_t) = 0.$$  

This trimming procedure is related to the Winsorized Mean estimator for censored models with only exogenous regressors proposed by Lee (1992). In the Winsorized Mean estimation, an arbitrary constant (instead of $x_t'\beta_0$) is introduced to perform the trimming in the SCLS, in order to allow for some asymmetry in the error distribution.\footnote{In Lee (1992), the arbitrary constant cannot be determined in the estimation.} The SCLS procedure requires that $u_t^*$ is symmetric up to $\pm x_t'\beta_0$, while the moment condition (2.6) only requires it be symmetric up to $\pm z_t'\beta_0$. Thus, given the condition (2.4), the symmetry assumed in equation (2.6) is weaker than that in the SCLS.

The condition $z_t'\beta_0 > 0$ in equation (2.6) follows the condition $x_t'\beta_0 > 0$ in SCLS. As discussed in Powell (1986), for the SCLS, if $x_t'\beta_0 \leq 0$, then most of the distribution of $u_t^*$ has been censored to $-x_t'\beta_0$. Symmetrically censoring the upper tail of $u_t$ will result in censoring all values of $u_t$ to $x_t'\beta_0$, thus it yields no information concerning the unknown $\beta_0$. Therefore, observations with $x_t'\beta_0 \leq 0$ are deleted from the sample. Similarly, to ensure the unique identifiability, $z_t'\beta_0$ is required to be positive for a substantial proportion of the sample. In practice, because every $y_t^*$ itself is censored to be above zero, it is reasonable for $x_t'\beta_0$ to be non-negative for a positive fraction of individuals in the sample. Generally the value of $z_t'\beta_0$ should be close to the values of $x_t'\beta_0$ because there are very few endogenous regressors in $x$ for most empirical models. Hence, the condition that $z_t'\beta_0$ is positive for a proportion of the sample will likely to be satisfied.

The requirement that $z_t'\beta_0 > 0$ for a proportion of the sample and that $z_t'\beta_0 \leq x_t'\beta_0$ for all $t$ puts a bound for the instrument. When there is one endogenous regressor, the condition (2.3) essentially requires that the value of instrument is smaller or larger than the endogenous regressor
Such instruments are not uncommon in empirical work. For example, Butcher and Case (1994) use the presence of any sister within a family as an instrumental variable for years of schooling of female workers, Levitt (1997) uses the timing of mayoral and gubernatorial election as instruments for size of the police force in estimating the effects of police on city crime rates. Card (1995) employs a dummy variable that indicates whether a man grew up in the vicinity of a four-year college as an instrumental variable for years of schooling. In these cases, the instrument takes the values 0 or 1, and is certainly smaller than the corresponding endogenous regressor. Thus, the condition (2.3) is satisfied. In many time series applications, lagged values are used as instruments; and lagged values are smaller than the current values in some cases, for example, lagged consumption or GDP. Therefore, they also meet the requirement.

The moment condition (2.6) can generate the following $l$ unconditional ones:

\[(2.7) \quad E[1(z^l_0) > 0] \cdot \min \{\max \{u^*, -z^l_0\}, z^l_0\} \cdot w_i = 0.6\]

Since $z$ is only a subset of instrument $w$, the above moment condition has the advantage of accommodating additional instruments. Thus, all information provided by the full set of instruments $w$ can be used to improve the efficiency, while only some of the instruments in $z$ need to satisfy the condition (2.3). Moreover, with over-identifying restrictions, it is possible to test for the model specification or the validity of some instruments. Since $-z^l_0$ falls into the uncensored region of $u^*$, condition (2.7) is equivalent to:

\[(2.8) \quad E[1(z^l_0 > 0) \cdot \min \{\max \{u, -z^l_0\}, z^l_0\} \cdot w_i] = 0,\]

where $u = \max \{u^*, -z^l_0\}$, and $u^* = y - x^l_0$.

The sample counterpart of the above population orthogonality condition is defined as:

---

5 There are considerable flexibilities in manipulating an instrument to satisfy this requirement, for example, the negative of an instrument is still a good instrument, and the same is true when an instrument plus a constant. Hence, the condition (2.3) is not as restrictive as it first appears.

6 The symbol “$1(A)$” denotes the indicator function of the event “$A$”, i.e., it is a function which takes the value one if $A$ is true and is zero otherwise.
\[
\frac{1}{n} \sum_{t=1}^{n} \left[ 1(z_t' \beta > 0) \cdot \min \{ \max \{ y_t - x_t' \beta, -z_t' \beta \}, z_t' \beta \} \cdot w_t \right] = 0.
\]

It is difficult to ensure identification using only these moment conditions, since multiple solutions to the moment condition exist. The usual approach is to integrate back the moment function to get a global minimand to identify the true parameter. In the SCLS, the criterion function is obtained by integrating the just-identified sample orthogonal system, and the observations with \( x_t' \beta_0 \leq 0 \) are “penalized” by the amount \( y_t^2/2 \) in the criterion function. In this construction, the SCLS criterion function avoids the multiplicity of minimum solutions.

However, the SCLS structure cannot be applied here because the trimming is performed by \( z \) instead of \( x \), and the number of instruments may be larger than that of endogenous regressors. Nevertheless, following Hansen (1982), a quadratic form GMM criterion function based on the moment condition can be constructed. By minimizing the GMM type criterion function, the \( k \) linear combinations of the sample orthogonality conditions are made as close as possible to zero. This gives \( k \) parameter-defining mapping equations, which can be solved for \( \beta \).

Since the GMM estimator is defined on the symmetric censored moment functions, this estimator will be called symmetrically censored GMM estimator (SCGMM). It is clear that the SCGMM estimator includes the SCLS as a special case for \( w_t = z_t = x_t \).

Therefore, the GMM criterion function is defined as

\[
Q_n(\beta) = F_n(\beta)' A_n F_n(\beta), \tag{2.10}
\]

where \( F_n(\beta) = \frac{1}{n} \sum_{t=1}^{n} f_t(\beta) \) and \( f_t(\beta) = 1(z_t' \beta > 0) \cdot \min \{ \max \{ y_t - x_t' \beta, -z_t' \beta \}, z_t' \beta \} \cdot w_t \), and \( A_n \) is a sequence of arbitrary, symmetric, positive definite, and possibly data-dependent weighting matrices. Under suitable regularity conditions, it can be shown that the GMM criterion function will converge to the following limiting function,

\[
Q_0(\beta) = M(\beta)' A_n^* M(\beta), \tag{2.11}
\]
where \( M(\beta) = \frac{1}{n} \sum_{t=1}^{n} m_t(\beta) \) and \( m_t(\beta) = \mathbb{E}[1(z_t' \beta > 0) \cdot \min\{\max\{ y_t - x_t' \beta, -z_t' \beta\}, z_t' \beta\} \cdot w_t] \), and \( A_n^* \) is an \( O(1) \) sequence of non-stochastic, symmetric, uniformly positive definite matrices. And \( m_t(\beta) \) and \( M(\beta) \) are the expectations of \( f_t(\beta) \) and \( F_n(\beta) \), respectively.

For the limiting criterion function \( Q_n(\beta) \), given the moment condition (2.8) and positive definiteness of \( A_n^* \), the minimum zero will be achieved at the true value \( \beta_0 \). However, a particular problem for identification is that the minimum can also be achieved at other \( \beta \)'s that makes \( z_t' \beta \leq 0 \) for all \( t \). For example, \( \beta = 0 \) will always achieve the minimum of \( Q_n(\beta) \). Given the moment condition, it is difficult to construct a criterion function that can exclude all unwanted roots to identify the true parameter. The SCGMM estimator, instead, is defined as a constrained estimator. In particular, the parameter space is constrained so that all unwanted roots are excluded. This approach is different from a usual extreme estimator, in which the parameter space is not explicitly constrained. In fact, a constrained optimization is more desirable if the researcher believes that the true parameter lies in a proper constrained subset (Amemiya 1985).

Therefore, the SCGMM estimator \( \hat{\beta}_n \) of \( \beta_0 \) is defined as the value of \( \beta \) minimizing \( Q_n(\beta) \) over a constrained subset of the parameter space. That is

\[
Q_n(\hat{\beta}_n) = \inf_{\beta \in \Theta_n} Q_n(\beta)
\]

s. t. \( \Theta_n = \{ \beta \in \Theta : \frac{1}{n} \sum_{t=1}^{n} 1(y_t > 0) z_t' \beta > 0 \} \),

whenever \( n > n_0 \), for some positive \( n_0 \).\(^7\)

The compact parameter space \( \Theta_n \) is a subset of \( \Theta \), and it can be viewed as a sufficiently large ball that contains \( \beta_0 \), in which the sample average of \( z_t' \beta \) for positive \( y_t \) is positive. Hence, in the parameter space \( \Theta_n \), all unwanted roots such as \( \beta = 0 \) are ruled out, and only the true value \( \beta_0 \)

\(^7\) I am grateful for the suggestion from a referee to specify the constraint this way.
satisfies the moment condition and minimizes the limiting SCGMM criterion function. Given the regularity condition discussed in next section, the true parameter can be identified in $\Theta_n$. Clearly, $\Theta_n$ is random set depending on the realization of $z_t$ and $y_t$. Thus, the SCGMM estimator becomes an extreme estimator constrained to the random set that converges in a certain probabilistic sense to a fixed set. The random set poses no particular problem for proving consistency and asymptotic normality. In practice, to find the minimum in a constrained parameter space, a mathematical programming method is needed in estimation. Given the availability of mathematical programming packages in computer software, the computation does not pose a problem.

### III. Consistency of the SCGMM Estimator

In this section, the consistency of the SCGMM estimator is established for processes that are near epoch dependent (NED) on a mixing process. The following definitions are based on Gallant and White (1988):

**Definition 3.1**: Let $\{V_t: \Omega \rightarrow \mathbb{R}^b\}$ be a sequence of random variables on a probability space $(\Omega, F, P)$, where $b \in \mathbb{N}$. The triangular array $\{Z_{nt}: t = 1, \ldots, n, n \in \mathbb{N}\}$ of random variables on $(\Omega, F, P)$ is near epoch dependent (NED) on $\{V_t\}$ if $E[|Z_{nt}|^2] < \infty$ for any $t \leq n, n \in \mathbb{N}$ and $\nu(m) \rightarrow 0$ as $m \rightarrow \infty$, where $\nu(m) \equiv \sup_n \sup_{t \in \{1, \ldots, n\}} \{E[Z_{nt} - E^{t+m}_{t-m}(Z_{nt})]^2: t \in \{1, \ldots, n\}, n \in \mathbb{N}\}$, and $E^{t+m}_{t-m}(\cdot) = E(\cdot | F^{t+m}_{t-m})$, $F^{t+m}_{t-m} = \sigma(V_{t+m}, \ldots, V_{t+m})$, and $L_p$ norm $\|Z\|_p = E|Z|^p$.

**Definition 3.2**: In definition 3.1, suppose $\nu(m) = O(m^\lambda)$ for all $\lambda < -a$. Then $\nu(m)$ is said to be of size $-a$. The size for mixing is defined similarly.

In the definition for NED, $\nu(m)$ is basically the root of the worst mean square forecast error when $Z_{nt}$ is predicted by $E^{t+m}_{t-m}(Z_{nt})$; and $\nu(m)$ will never increase as $m$ increases for all $t$, as more and more information is used for forecasting. If $\nu(m)$ tends to zero at an appropriate rate
uniformly in $t$, then $Z_n$ depends essentially on the recent epoch of $\{V_t\}$. If $Z_n$ is independent, $m$-dependent, or mixing, then it is trivially near epoch dependent of any size.

Let the $L_p$ norm of the random variables $u_t^*, x_t$, and $w_t$ be denoted as $\|u_t^*\|_p = E^{1/p} |u_t|^p$, $\|x_t\|_p = E^{1/p} |x_t|^p$, and $\|w_t\|_p = E^{1/p} |w_t|^p$, where $\|x_t\|$ and $\|w_t\|$ are Euclidean norms. The following assumptions are needed for consistency.

**Assumption E:** The error terms $u_t^*$ are continuously and symmetrically distributed about zero conditionally on $w_t$ with continuous density function $g_t(\lambda)$, where $g_t(\lambda)=g_t(-\lambda)$, $g_t(\lambda)<L_0$, and $g_t(\lambda)>\xi_0$ in the neighborhood of zero uniformly in $t$ for some $L_0>0$, $\xi_0>0$.

Assumption E implies the condition $E(u_t^* | w_t)=0$ and $m_t(\beta_0)=0$.

**Assumption MX:** $\{x_t, w_t\}$ is a mixing sequence such that either $\phi_m$ is of size $-r/(2r-2)$, $r\geq 2$ or $\alpha_m$ is of size $-r/(r-2)$ with $r>2$.

**Assumption NE:** The elements of sequence $\{u_t^*\}$ are near epoch dependent on $\{x_{1t}, v_t\}$ of size $-(r-1)/(r-2)$ for $r>2$, where $\{v_t\}$ is a mixing sequence such that either $\phi_m$ is of size $-r/(2r-2)$, $r\geq 2$ or $\alpha_m$ is of size $-r/(r-2)$ with $r>2$.

**Assumption DM:** The random variables $u_t^*, x_t$, and $w_t$ are $2r$-integrable uniformly in $t$, for $t = 1, 2..., \text{ and } r>2$, that is, their $L_{2r}$ norms are bounded for all $t$, $\|u_t^*\|_{2r} \leq \Delta < \infty$, $\|x_t\|_{2r} \leq \Delta < \infty$, and $\|w_t\|_{2r} \leq \Delta < \infty$.

The mixing assumption of $\{x_t, w_t\}$ generalizes the independence assumption in SCLS. If the instrument vector $w_t$ includes interaction terms or functions of some exogenous elements in $x_t$, the mixing property of these terms is implied by the mixing property of $x_t$, because a function of mixing processes is mixing. It will suffice to assume just $\alpha$-mixing (strong mixing), because $\phi$-mixing (uniform mixing) implies $\alpha$-mixing. The domination condition places bounds on some
moments greater than the fourth, which is a stronger requirement. However, this will allow for much greater dependence and fairly arbitrary heterogeneity.

It would be technically easier if the disturbance term $u_t^*$ is also a mixing process, which still extends the GMM from stationary sequences to heterogeneous sequences and makes the SCGMM applicable to models with dependent observations. For many cases, the mixing assumption is sufficient because if $u_t^*$ only depends on a finite number of lagged values of some mixing processes, $u_t^*$ itself is still mixing. However, the mixing assumption has a drawback. A function of a mixing sequence (or even an independent sequence) that depends on an infinite number of lags of the sequences is not generally mixing. The property of $\alpha$-mixing ($\phi$-mixing) is not necessarily preserved under transformations that involve the infinite past (Andrews, 1984). For example, even if $u_t^*$ is a simple AR(1) process on an underlying process $v_t$ and $v_t$ is mixing or even independent, $u_t^*$ itself can fail to be either $\alpha$-mixing or $\phi$-mixing.

Assumption NE adopts a more general dependence structure, near epoch dependence. Based on the definition, a process may depend on the entire history of other processes, but if $\nu(m)$ tends to zero at an appropriate rate, then it depends essentially on the recent epoch and does not depend “too much” on the distant past or future. Therefore, Assumption NE allows $u_t^*$ to depend on the infinite past and/or future of the underlying processes, provided the extent of dependence is controlled appropriately. Since some elements of $x$ are endogenous, $u_t^*$ is allowed to depend on these endogenous elements in the NED structure. It can be shown that, if $u_t^*$ is an ARMA process of finite order with zeros lying outside the unit circle, or if $u_t^*$ is an infinite moving average process, $u_t^*$ is NED. If $u_t^*$ is a mixing process, it is trivially NED. In addition, we need not impose stationarity on $u_t^*$ but instead may allow a substantial amount of heterogeneity. Since near epoch dependent functions of mixing processes are mixingales, asymptotic theories for mixingales can be applied (see Gallant and White 1988 for further discussions). Therefore, the
NED assumption is general enough for the SCGMM estimator to apply to models with a variety of dependence structures.

**Assumption P:** The parameter space $\Theta$ is a compact subset of $\mathbb{R}^k$, and the true parameter vector $\beta_0$ is an interior point of $\Theta$.

**Assumption ID:** i) $z_t$ is a subset of $w_t$ such that $z_t = (w_{1t}, x_{2t})$, and $x_t' \beta_0 \geq z_t' \beta_0$, for all $t$.

ii) The matrix $N_n = \frac{1}{n} \sum_{t=1}^{n} E[z_t' \beta_0 > \varepsilon_0] w_t w_t'$ is positive definite uniformly in $n$, whenever $n > n_0$, for some positive $n_0$ and $\varepsilon_0$.

iii) $\frac{1}{n} \sum_{t=1}^{n} E(z_t' \beta_0 | y_t > 0) > 0$, uniformly in $n$, whenever $n > n_0$, for some positive $n_0$.

iv) The rank of the matrix $\frac{1}{n} \sum_{t=1}^{n} E[w_t x_t']$ is $k$.

Assumption ID is essential for identification. Condition i) is the bound condition for a subset of the instrument discussed in the previous section, and it essentially requires that $z_t' \beta_0$ fall into the region of $u_t$ that is not censored by $-x_t' \beta_0$. Condition ii) imposes a lower bound to $z_t' \beta_0$, and requires $z_t' \beta_0$ to be non-negative for a substantial proportion of the sample. It also rules out the situations that some elements of $w_t$ are perfect collinear.

Condition iii) requires that the average value of $z_t' \beta_0$ for positive $y_t^*$ to be positive, and it ensures the convergence of the compact subset $\Theta_n$. Based on condition ii), $z_t' \beta_0$ is positive for a substantial proportion of the sample. Since the “true” dependent variable $y_t^*$ is censored to be above zero, $x_t' \beta_0$ is more likely to be positive if $y_t^*$ itself is positive. Hence, $z_t' \beta_0$ is also more likely to be positive when $y_t^* > 0$ than when $y_t^* < 0$. This assumption is especially convenient for implementing the SCGMM estimator because the constraint becomes linear. Condition iv) is the usual rank condition for IV estimation, and requires $w_t$ and $x_t$ be sufficiently linearly related.

**Assumption W:** The sequence of weighting matrices $A_n$ is symmetric, positive definite,
and $A_n \overset{p}{\longrightarrow} A_n^*$, where $A_n^*$ is nonstochastic, symmetric, and uniformly positive definite.$^8$

The above assumptions permit application of Gallant and White’s (1988) version of the uniform law of large numbers (Theorem 3.18) based on near epoch dependence and the Lipschitz condition. The Lipschitz condition is defined below.

**Definition 3.3:** Let $(\Omega, F, P)$ be a probability space and $(\Theta, \rho)$ be an Euclidean space. The sequence $\{q_t: \Omega \times \Theta \rightarrow \mathbb{R}\}$ is defined to be almost surely Lipschitz-L1 on $\Theta$ if and only if for each $\theta$ in $\Theta$, $q_t(\cdot, \theta)$ is measurable-F, $t=1, 2, \ldots$, and for any $\theta, \theta_0$ in $\Theta$ there exist a function $L_t^0: \Omega \rightarrow \mathbb{R}^+$ measurable-F/B(\mathbb{R}^+) such that

$$\frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[L_t^0] \text{ is } O(1), \quad \text{and}$$

$$\left| q_t(\theta) - q_t(\theta_0) \right| \leq L_t^0 \| \theta - \theta_0 \|, \quad t = 1, 2, \ldots, \text{ a.s.}$$

The following Lemmas are needed to establish the large sample properties.

**Lemma 3.1:** Given assumptions DM and P, the elements of $f_t(\beta)$ are almost surely Lipschitz-L1 on $\Theta$.

**Lemma 3.2:** Given assumptions DM, MX, NE, and P, the elements of $f_t(\beta)$ are near epoch dependent on $\{w_t, x_t, v_t\}$ of size $-1/2$ uniformly on $(\Theta, \rho)$, where $\rho$ is Euclidean norm on $\mathbb{R}^k$.

**Lemma 3.3:** Given assumption E, P, and ID, there exists a sequence of subset $\overline{\Theta}_n$ of $\Theta$, such that $\Theta_n$ contains $\beta_0$, and for every $\beta$ in $\overline{\Theta}_n$, $\left[ \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}(z_i^t | y_i > 0) \right] > 0$ uniformly in $n$ and uniformly in $\beta$, whenever $n > n_0$ for some positive $n_0$, and the moment condition $M(\beta) = 0$ is only satisfied at $\beta_0$ over $\overline{\Theta}_n$.

---

$^8$ $A_n^*$ is used here because, unlike that in Hansen (1982), $A_n^*$ is not limited to converge to a constant limit (see Bates and White, 1985).
Since the random set $\Theta_n$ will converge to the fixed set $\Theta^*$, based on Lemma 3.3, the limiting SCGMM criterion function $Q_0(\beta)$ has identifiably unique minimizer $\beta_0$ on $\Theta^*$. With these conditions, the following result can be established:

**THEOREM 3.4:** For the censored regression model (2.1), under assumptions E, MX, NE, DM, P, ID and W, there exists a measurable function $\hat{\beta}_n : \Omega \rightarrow \Theta_n$, such that

$$Q_n(\beta) = \inf_{\beta \in \Theta_n} Q_n(\beta) ; \text{ and the SCGMM estimator } \hat{\beta}_n \text{ is weakly consistent, } \hat{\beta}_n \xrightarrow{p} \beta_0. \quad (9)$$

**IV Asymptotic Normality of the SCGMM Estimator**

The asymptotic distribution of the SCGMM estimator is established through the empirical process method via stochastic equicontinuity. There are primitive conditions available to show stochastic equicontinuity. If the empirical moment function $f_t(\beta)$ is independent or m-dependent, the corresponding empirical process can be shown to be stochastically equicontinuous via symmetrization because $f_t(\beta)$ belongs to the type I class of functions which satisfy Pollard’s entropy condition (see Andrews 1994a). If $f_t(\beta)$ is mixing, stochastic equicontinuity can be shown via bracketing conditions (Andrews, 1993).

However, as $f_t(\beta)$ is near epoch dependent here, the above primitive conditions are not applicable. Nevertheless, since $f_t(\beta)$ is Lipschitz at $\beta_0$ and differentiable at the true parameter $\beta_0$ with probability one, a weaker stochastic equicontinuity condition can be established following Newey and McFadden (1994).\(^{10}\) The following assumptions are needed for the asymptotic distribution of the SCGMM estimator.

---

\(^9\) It could be strongly consistent if the weighting matrix $A_n$ could converge almost surely. However, as discussed in section V, the optimal weighting matrix only converges in probability.

\(^{10}\) As discussed on the Appendix, the stochastic equicontinuity condition for Lipschitz moment functions defined in Theorem 7.2 and 7.3 in Newey and McFadden (1994) can be extended to the NED case.
Assumption NE’: The disturbance process \( \{u_t^*\} \) is near epoch dependent with respect to \( \{x_{1t}, v_t\} \) of size \(-2(r-1)/(r-2)\) for some \( r > 2 \).

Assumption MX’: The process \( \{x_t, w_t, v_t\} \) is a mixing sequence such that either \( \phi_m \) is of size \(-r/(r-1)\), \( r \geq 2 \) or \( \alpha_m \) is of size \(-2r/(r-2)\) with \( r > 2 \).

Assumptions NE’ and MX’ strengthen assumptions NE and MX by increasing their sizes.

Define \( B_n^0 = \text{Var} (\sqrt{n} F_n(\beta_0)) \), and \( G_n^0 = \frac{1}{n} \sum_{t=1}^{n} \text{E}[1(-z_t^* \beta_0 < u_t^* < z_t^* \beta_0)(-w_t x_t')] \).

Assumption PD: The sequence \( \{B_n^0\} \) is uniformly positively definite.

In order to establish the asymptotic distribution, the following central limit theorem (Wooldridge and White 1988) for NED functions of a mixing process is used.

**Central Limit Theorem:** Let \( \{Z_{nt}\} \) be a double array such that \( \|Z_{nt}\| \leq \Delta < \infty \) for some \( r > 2 \), \( E(Z_{nt}) = 0 \), \( n, t = 1, 2, \ldots \), and \( \{Z_{nt}\} \) is near epoch dependent on \( \{V_t\} \) of size \(-1\), where \( \{V_t\} \) is a mixing process with \( \phi_m \) is of size \(-r/(r-2)\) or \( \alpha_m \) is of size \(-2r/(r-2)\). Define \( v_n^2 \equiv \text{Var} \left( \sum_{t=1}^{n} Z_{nt} \right) \), and suppose that \( v_n^{-2} \) is \( O(n^{-1}) \). Then \( v_n^{-1} \left( \sum_{t=1}^{n} Z_{nt} \right) \xrightarrow{d} N(0,1) \).

The following lemmas give the limiting distribution of \( F_n(\beta_0) \) and the stochastic equicontinuity condition. The stochastic equicontinuity in Lemma 4.2 is a weaker result than that in Andrews (1994a) because of the denominator term.

**Lemma 4.1:** Given assumptions E, DM, P, ID, MX’, NE’, and PD, then

\[
(B_n^0)^{-1/2} \sqrt{n} F_n(\beta_0) \xrightarrow{d} N(0, I).
\]

**Lemma 4.2:** Given assumptions E, DM, P, ID, and MX’, then for any \( \delta_n \to 0 \),

\[
\sup_{\|\beta - \beta_0\| \leq \delta_n} \sqrt{n} \left\| F_n(\beta) - F_n(\beta_0) - M(\beta) \right\| / \left[ 1 + \sqrt{n} \|\beta - \beta_0\| \right] \xrightarrow{p} 0.
\]
**THEOREM 4.3**: Given assumptions E, DM, P, ID, MX’, NE’, and PD, if $\hat{\beta}_n$ is a consistent estimator of $\beta_0$, then $(G_n^0 A_n^* B_n^0 A_n^* G_n^0)^{-1/2}(G_n^0 A_n^* G_n^0) \sqrt{n} (\hat{\beta}_n - \beta_0) \xrightarrow{d} N(0, I)$.

If $G_n^0, A_n^*$, and $B_n^0$ have limits $G^0, A^*$, and $B^0$, respectively, the result can be written in the following form: $\sqrt{n} (\hat{\beta}_n - \beta_0) \xrightarrow{d} N[0, (G^0 A^* G^0)^{-1} G^0 A^* B^0 A^* G^0 (G^0 A^* G^0)^{-1}]$.

V. Consistent Estimator of the Asymptotic Covariance Matrices

In order to use the asymptotic normality of $\hat{\beta}_n$ to conduct hypothesis tests for the parameter vector $\beta_0$, consistent estimators of the asymptotic covariance matrices must be derived. In addition, a consistent estimator of the asymptotic covariance matrix also provides the optimal weighting matrix to get a more efficient SCGMM estimator.

Based on Theorem 4.3, to estimate the asymptotic covariance matrices, both $G_n^0$ and $B_n^0$ must be estimated. A “natural” estimator of $G_n^0$ is

$$
\hat{G}_n (\hat{\beta}_n) = \frac{1}{n} \sum_{i=1}^{n} 1(-z_i^{*'} \hat{\beta}_{n} < u_i^{*'} < z_i^{*'} \hat{\beta}_n) \cdot (-w_i x_i'),
$$

where $u_i = y_i - x_i^{*'} \hat{\beta}_n$. It can be shown that $\hat{G}_n (\hat{\beta}_n)$ is a consistent estimator of $G_n^0$.

In the presence of arbitrary forms of heteroskedasticity and autocorrelation, $B_n^0$ can be written as $B_n^0 = \sum_{j=-n+1}^{n-1} \Gamma (j)$, where

$$
\Gamma (j) = \begin{cases} 
\frac{1}{n} \sum_{i=j+1}^{n} E[f(\beta_0) \cdot f'(\beta_0)] & \text{for } j \geq 0 \\
\frac{1}{n} \sum_{i=1}^{n} E[f(\beta_0) \cdot f'(\beta_0)] & \text{for } j < 0.
\end{cases}
$$

In the simplest case, such for an independent sequence, $B_n^0$ equals $\Gamma (0)$, where
\[ \Gamma(0) = \frac{1}{n} \sum_{t=1}^{n} \text{E} [1(z_t' \beta_0 > 0) \cdot (\min\{(u_t)^2, (z_t' \hat{\beta}_0)^2\}) \cdot w_tw_t'] \].

A consistent estimator for \( \Gamma(0) \) is:

\[ \hat{\Gamma}(0) = \frac{1}{n} \sum_{t=1}^{n} 1(z_t' \hat{\beta}_0 > 0) \cdot (\min\{(u_t)^2, (z_t' \hat{\beta}_0)^2\}) \cdot w_tw_t' \], where \( u_t = y_t - x_t' \hat{\beta}_0 \).

If observations are m-dependent and m is a known finite integer, a consistent estimator for \( B_n^0 \) is:

\[ \hat{B}_n^m = \hat{\Gamma}(0) + \sum_{j=1}^{m} \left[ \hat{\Gamma}(j) + \hat{\Gamma}(j)' \right] \], where we have used the fact that \( \hat{\Gamma}(-j) = \hat{\Gamma}(j)' \), and \( \hat{\Gamma}(j) = \frac{1}{n} \sum_{t=j+1}^{n} [f(\hat{\beta}_j) \cdot f'(\hat{\beta}_j)] \), for \( j \geq 0 \).

In a general case, the autocorrelation among elements of \( f(\beta) \) may not equal zero after a finite number of lags. Nevertheless, it is still possible to obtain useful estimators by using the near epoch dependence and mixing conditions. In this case, we can define the heteroskedasticity and autocorrelation consistent (HAC) covariance matrix estimator for \( B_n^0 \) as:

\[ \hat{B}_n(p_n) = \hat{\Gamma}(0) + \sum_{j=1}^{p_n} \left( 1 - \frac{j}{p_n + 1} \right) \cdot \left[ \hat{\Gamma}(j) + \hat{\Gamma}(j)' \right] \],

where the weight \( 1 - \frac{j}{p_n + 1} \) is suggested by Newey and West (1987) and guarantees the positive semi-definiteness of \( \hat{B}_n \). It is clear that as \( j \) increases, the weights decrease. For the choice of the lag truncation parameter \( p_n \), it is necessary to let \( p_n \) go to infinity at some suitable rate. The rate given in the following assumption also follows Newey and West (1987).

**Assumption TL:** \( \{p_n\} \) is a sequence of integers such that \( p_n \to \infty \) as \( n \to \infty \) and \( p_n = O(n^{1/4}) \).

**Assumption DM':** The random variables \( u_t^*, x_t, \) and \( w_t \) are \( 4r \)-integrable uniformly in \( t \), for \( r \geq 2 \), \( t = 1, 2, \ldots \).
**Assumption NE”**: The disturbance process \( \{u_t^*\} \) is near epoch dependent with respect to \( \{x_{1t}, v_t\} \) of size \(-4(r-1)^2/(r-2)^2\) for some \( r>2 \).

**THEOREM 5.1**: Given assumptions E, P, ID, MX’, TL, DM’, and NE”, if \( \hat{\beta}_n \) is a consistent estimator of \( \beta_0 \), then the estimators \( \hat{G}_n(\hat{\beta}_n) \) and \( \hat{B}_n(p_n) \) defined above are weakly consistent, \( \hat{G}_n(\hat{\beta}_n) \xrightarrow{p} G_0 \) and \( \hat{B}_n(p_n) \xrightarrow{p} B_0 \).

Furthermore, if the weighting matrix is chosen as \( \hat{B}_n(p_n) \), the SCGMM estimator will be relatively more efficient. Based on Hansen (1982), the weighting matrix \( \hat{B}_n(p_n) \) is called the optimal weighting matrix. It can be shown that the asymptotic covariance matrix based on the optimal weighting matrix is smaller than that based on other weighting matrix.

**Corollary 5.2**: If the weighting matrix is \( \hat{B}_n(p_n) \), then the SCGMM estimator \( \hat{\beta}^* \) has a smaller variance than that based on other weighting matrices with different limits, and the asymptotical distribution becomes \( \left[ G_0^0(B_0^{-1} G_0^0)^{1/2} \sqrt{n} (\hat{\beta}^* - \beta_0) \right] \xrightarrow{d} \mathcal{N}(0, I) \).

In practice, two steps are needed in order to obtain the optimal weighting matrix because it is necessary to have a preliminary consistent estimate of the parameter before \( \hat{B}_n(p_n) \) can be calculated. The preliminary consistent estimate can be obtained by using any arbitrary weighting matrix such as the identity matrix. If the model is just identified, the choice of weighting matrices is irrelevant, and the asymptotic distribution has the same simple expression as that with the optimal weighting matrix.

With the SCGMM estimator, several specification tests are possible. A test of over-identifying restrictions can follow Hansen (1982). It can be shown that \( nF_d(\hat{\beta}_n^*) \xrightarrow{d} \mathcal{B}_d(p_n) F_d(\hat{\beta}_n^*) \) converges in distribution to a chi-square distributed random variable with \( l-k \) degrees of freedom. A related test on the validity of the instruments is also possible based on Newey (1985).
by comparing the estimates using all instruments and using a subset of valid instruments to construct a Hausman type test statistic.

VI. Simulation Results

In this section, a small scale simulation study of a simple linear regression model is conducted. The dependent variable y is generated from the equation $y_t^* = \alpha + \beta x_t + u_t$ and $y_t = \max(0, y_t^*)$. In order to focus on the performance of the trimming procedure conducted by the instruments, the simulation model is generated by i.i.d. processes. The parameter vector has an intercept $\alpha$ and a slope $\beta$. In order to create a correlation between them, $x_t$ and $u_t$ are generated as $u_t = \varepsilon_t + e_t$ and $x_t = \pi + \lambda e_t + v_t$, where $\varepsilon_t$ is generated from i.i.d. standard normal distribution $N(0,1)$, and $v_t$ assumes evenly-spaced values in the interval $[-b, b]$ so that the variance $\text{Var}(v_t) = 1$ (i.e., $b \approx 1.7$), and $e_t$ takes on the repeated sequence $\{-1, 1, -1, 1, \ldots\}$ and $\text{Var}(e_t) = 1$. In this case, $E(e_t) = 0$ and $E(v_t) = 0$, and the correlation between $v_t$, $e_t$, and $e_t$ is zero.

These specifications also imply that $E(u_t) = 0$ and the regression error $u_t$ is symmetrically distributed about zero. The instrument $z_t$ is generated as $z_t = \delta + v_t$. The condition that $x_t'\beta \geq z_t'\beta$ for all $t$ requires that $\pi \geq \delta - \lambda e_t$. Given the values of $e_t$, this condition requires $\pi \geq \delta + \lambda$. In this case, the correlation between $x$ and $u$ is $\rho_{ux} = \frac{\lambda}{\sqrt{(\chi^2 + 1)^3}}$. The correlation between $x$ and $z$ is $\rho_{xz} = \frac{1}{\sqrt{(\chi^2 + 1)^3}}$.

As in any IV estimation, the instrument should be correlated with the endogenous regressor, and this correlation affects the asymptotic variance of the SCGMM estimator. Both correlation coefficients are determined by parameter $\lambda$. Clearly, the regression error $u$ is not normally distributed, and thus the Tobit procedure cannot be used. Moreover, since the regressor $x$ is correlated with the error $u$, the SCLS is not consistent either.

The implementation of the SCGMM estimation requires minimization of the non-smooth criterion function $Q_n(\beta)$ over the parameter space $\Theta_n$. Usual algorithms based on unconstrained parameter spaces generally do not work and almost always result in solutions where $z_t'\beta \leq 0$ for all
Therefore, based on the constraint on the parameter space, a mathematical programming method is needed for the computation. The constraint requires that the average predicted values of \( z_t' \beta \) for observations with positive \( y_t \) are positive, and thus it becomes essentially linear. This condition is especially convenient for computation.

Additionally, the criterion function itself is not continuously differentiable, which causes problems for finding the optimal value in the mathematical programming process. Some approximations are used to smooth the criterion function. In the simulation, the weighting matrix used is the identity matrix. Because the model is just-identified, the choice of weighting matrix is immaterial.

In the basic scenario, the parameters are defined as \( \alpha = 1, \beta = 1, \lambda = 1, \delta = 0 \). In this case, the correlation between \( x \) and \( u \) is 0.5, and between \( x \) and \( z \) is 0.71, and the censoring proportion of \( y_t^* \) is 25%. In addition, about 78% of the values of \( (\alpha + \beta z_t) \) are positive. As indicated in the sections above, values of negative \( (\alpha + \beta z_t) \) are dropped in the estimation. For each simulation design, the “true” parameter values are reported along with the sample mean, standard deviation (SD), root mean squared error (RMSE), median, and median absolute errors (MAE). For comparison, the results based on Powell’s SCLS estimator are also reported. For each design, the simulation is repeated for 400 times. The simulation results are summarized in Table 1.

---

11. In the simulation, the software package GAMS (the General Algebraic Modeling System) is used for computation via its mathematical programming solvers MINOS. OLS estimates based on the censored value \( y_t \) are used as starting values. Other software programs such as GAUSS also have mathematical programming packages.

12. In the implementation, a small positive constant is used as the lower bound for the constraint instead of using 0, because the mathematical programming solver allows the constraint to reach the bound. The results are not sensitive to the choice of this constant.

13. For example, the following approximations are used: \( \max(x, 0) = 0.5 \cdot \left( \frac{\sqrt{x^2 + \xi^2} + x}{2x + \xi} \right) \), where \( \xi \) is a small constant.

14. The SCLS estimator is calculated using (2.13) as a recursion formula (Powell, 1986), also with OLS estimates as starting values.

15. Sometimes, the solver fails to find the interior optimal solution (for example, the constraint reaches the lower bound). In this case the corresponding estimates are discarded. Such cases do not occur often. For example, in Design 1, among 400 repetitions, only one fails to find optimal solution; in Design 2, every repetition finds the optimal solution. When the degree of censoring increases, such cases seem to increase.
One feature of the tabulated results is the dramatic difference between the SCGMM estimates and the SCLS estimates. As we have shown, when regressors are correlated with the error terms, the SCLS estimation is inconsistent. This is evident in the table, and the SCGMM performs much better than does the SCLS. The results also show the finite sample bias of the SCGMM estimator, but it is small. The sample mean and median of the SCGMM estimator are very close to each other.

Throughout the table, the performance of the SCGMM improves as the sample size increases, measured by both SD and RMSE. When the sample size decreases by 50 percent (Design 2), the resulting standard deviation increases by about $\sqrt{2}$, which is in accord with the asymptotic theory. In Design 3 and 4, the intercept $\alpha$ is changed to zero. In this case, the censoring proportion of $y_i^*$ increases to about 38%, and the proportion of positive values of $(\alpha + \beta z_t)$ reduce to 50%. The increased censoring appears to have a strong effect on the performance of the SCGMM estimator. The mean bias increases, and so does the RMSE.

In the Design 5 and 6, $\lambda$ is changed to 0.6. This change increases the correlation between $x$ and $z$ to about 86% and reduces the correlation between $x$ and $u$ to about 37%. As expected, such a change improves the performance of the SCGMM estimator. The results show a reduced mean bias, and especially reduced standard deviation and RMSE for the intercept estimate.

Since the regression error consists of a standard normal random variable and a binary random variable (with the values -1 and 1), its distribution depends on these two random variables. In Design 7 and 8, the distribution of $\varepsilon$ is changed to $N(0, 3)$ to change the distribution of the regression error. The resulting new distribution has a larger variance of 4 (double the previous variance). As expected, the resulting SCGMM estimates show a relatively larger SD and RMSE.

It is impossible to completely characterize the finite sample behavior of the SCGMM estimator under the conditions imposed in the sections above. However, the simulation still
shows the relative performance for a few different cases characterized by different sample sizes, censoring proportions, correlations between the regressor, the regression error, and the instruments, and different distributions of the error term. In any case, the SCGMM estimator performs very well and much better than does the SCLS estimator.

**VII. Conclusion**

This paper proposes a semi-parametric SCGMM estimator for censored models. The estimator is robust to non-normality and heteroskedasticity of unknown forms for the distribution of the regression error, to the presence of endogenous explanatory variables, and to models with dependent observations. The SCGMM estimator is shown to be consistent and asymptotically normal. An HAC estimator of the covariance matrices is also provided. The SCGMM estimator also adopts a different approach to the identification of the true parameter by enforcing a constraint for the parameter space. The simulation results demonstrate that the SCGMM works very well, much better than does the SCLS estimator in the presence of endogenous regressors.

The SCGMM estimator needs a stronger requirement on the instruments than the usual GMM or IV estimation. However, as discussed, the requirement is not as restrictive as it appears in practice. In fact, the requirement that $z_t'\beta_0 \leq x_t'\beta_0$ for all $t$ can be relaxed. In particular, suppose there is only one endogenous regressor $x_1$ with the instrument is $z_1$, and if $x_1 \geq z_1$ for a substantial proportion in the population, then the condition $x_{1t}\beta_{01} \geq z_{1t}\beta_{01}$ holds for those realizations (suppose the true value $\beta_{01}$ is non-negative). Other realizations of $x_{1t}$ and $z_{1t}$ such that $x_{1t}\beta_{01} < z_{1t}\beta_{01}$ cannot help to identify the true value of $\beta_0$ and thus can be deleted. Therefore, in such a sample, the subsample with $x_1 \geq z_1$ can be used for estimation. As the sample size goes to infinity, the size of the subsample will also go to infinity, and then the true value $\beta_0$ can be identified. This is extremely useful when it is difficult to find instruments that satisfying condition (2.3) for all $t$.
The SCGMM estimator still requires that the error distribution be symmetric. A test on the symmetry assumption can be done based on Newey (1987), using some odd function of \( f(\beta) \) to construct additional moment conditions. In future research, it will be useful to relax the assumption of symmetry and to extend this approach to general sample selection where the censoring (selecting) index may be different from the dependent variable of the current regression equation.\(^{16}\)

\(^{16}\) Honoré and Powell (1994) propose a pairwise difference estimator that does not require the symmetry assumption for the error distribution, but the error terms must be an i.i.d. process and all regressors must be exogenous.
REFERENCES


Table 1  Simulation Results

<table>
<thead>
<tr>
<th>True</th>
<th>Mean</th>
<th>SD</th>
<th>RMSE</th>
<th>Median</th>
<th>MAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Design 1:  T=400, λ=1, censoring=25%, positive (α+βz)=78%, ρxz=0.71, ρxu=0.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCGMM</td>
<td>α 1</td>
<td>0.983</td>
<td>0.169</td>
<td>0.170</td>
<td>1.004</td>
</tr>
<tr>
<td></td>
<td>β 1</td>
<td>1.011</td>
<td>0.100</td>
<td>0.100</td>
<td>1.005</td>
</tr>
<tr>
<td>SCLS</td>
<td>α 1</td>
<td>0.520</td>
<td>0.097</td>
<td>0.489</td>
<td>0.528</td>
</tr>
<tr>
<td></td>
<td>β 1</td>
<td>1.498</td>
<td>0.053</td>
<td>0.501</td>
<td>1.498</td>
</tr>
<tr>
<td>Design 2:  T=200, λ=1, censoring=25%, positive (α+βz)=78%, ρxz=0.71, ρxu=0.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCGMM</td>
<td>α 1</td>
<td>0.932</td>
<td>0.245</td>
<td>0.255</td>
<td>0.958</td>
</tr>
<tr>
<td></td>
<td>β 1</td>
<td>1.039</td>
<td>0.137</td>
<td>0.143</td>
<td>1.026</td>
</tr>
<tr>
<td>SCLS</td>
<td>α 1</td>
<td>0.523</td>
<td>0.141</td>
<td>0.498</td>
<td>0.516</td>
</tr>
<tr>
<td></td>
<td>β 1</td>
<td>1.496</td>
<td>0.074</td>
<td>0.502</td>
<td>1.498</td>
</tr>
<tr>
<td>Design 3:  T=400, λ=1, censoring=38%, positive (α+βz)=50%, ρxz=0.71, ρxu=0.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCGMM</td>
<td>α 0</td>
<td>0.144</td>
<td>0.196</td>
<td>0.243</td>
<td>0.132</td>
</tr>
<tr>
<td></td>
<td>β 1</td>
<td>0.932</td>
<td>0.109</td>
<td>0.129</td>
<td>0.934</td>
</tr>
<tr>
<td>SCLS</td>
<td>α 0</td>
<td>-0.393</td>
<td>0.156</td>
<td>0.423</td>
<td>-0.380</td>
</tr>
<tr>
<td></td>
<td>β 1</td>
<td>1.465</td>
<td>0.073</td>
<td>0.470</td>
<td>1.461</td>
</tr>
<tr>
<td>Design 4:  T=200, λ=1, censoring=38%, positive (α+βz)=50%, ρxz=0.71, ρxu=0.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCGMM</td>
<td>α 0</td>
<td>0.189</td>
<td>0.219</td>
<td>0.290</td>
<td>0.200</td>
</tr>
<tr>
<td></td>
<td>β 1</td>
<td>0.918</td>
<td>0.132</td>
<td>0.155</td>
<td>0.913</td>
</tr>
<tr>
<td>SCLS</td>
<td>α 0</td>
<td>-0.386</td>
<td>0.216</td>
<td>0.442</td>
<td>-0.385</td>
</tr>
<tr>
<td></td>
<td>β 1</td>
<td>1.461</td>
<td>0.101</td>
<td>0.472</td>
<td>1.464</td>
</tr>
</tbody>
</table>
Table 1-continued: Simulation Results

<table>
<thead>
<tr>
<th>Design</th>
<th>True</th>
<th>Mean</th>
<th>SD</th>
<th>RMSE</th>
<th>Median</th>
<th>MAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>5: T=400, λ=0.6, censoring=25%, positive (α+βz)=78%, ρxz=0.86, ρxu=0.37</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCGMM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>α</td>
<td>1</td>
<td>0.990</td>
<td>0.132</td>
<td>0.133</td>
<td>1.007</td>
<td>0.104</td>
</tr>
<tr>
<td>β</td>
<td>1</td>
<td>1.010</td>
<td>0.100</td>
<td>0.100</td>
<td>1.005</td>
<td>0.080</td>
</tr>
<tr>
<td>SCLS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>α</td>
<td>1</td>
<td>0.680</td>
<td>0.150</td>
<td>0.353</td>
<td>0.691</td>
<td>0.118</td>
</tr>
<tr>
<td>β</td>
<td>1</td>
<td>1.462</td>
<td>0.099</td>
<td>0.472</td>
<td>1.456</td>
<td>0.077</td>
</tr>
<tr>
<td>6: T=200, λ=0.6, censoring=25%, positive (α+βz)=78%, ρxz=0.86, ρxu=0.37</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCGMM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>α</td>
<td>1</td>
<td>0.941</td>
<td>0.212</td>
<td>0.220</td>
<td>0.970</td>
<td>0.158</td>
</tr>
<tr>
<td>β</td>
<td>1</td>
<td>1.043</td>
<td>0.146</td>
<td>0.153</td>
<td>1.027</td>
<td>0.114</td>
</tr>
<tr>
<td>SCLS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>α</td>
<td>1</td>
<td>0.661</td>
<td>0.219</td>
<td>0.403</td>
<td>0.672</td>
<td>0.177</td>
</tr>
<tr>
<td>β</td>
<td>1</td>
<td>1.473</td>
<td>0.136</td>
<td>0.492</td>
<td>1.466</td>
<td>0.109</td>
</tr>
<tr>
<td>7: T=400, λ=1, censoring=30%, positive (α+βz)=78%, ρxz=0.71, ρxu=0.23, ε~N(0,3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCGMM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>α</td>
<td>1</td>
<td>1.026</td>
<td>0.371</td>
<td>0.372</td>
<td>1.106</td>
<td>0.292</td>
</tr>
<tr>
<td>β</td>
<td>1</td>
<td>0.991</td>
<td>0.241</td>
<td>0.241</td>
<td>0.956</td>
<td>0.192</td>
</tr>
<tr>
<td>SCLS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>α</td>
<td>1</td>
<td>0.560</td>
<td>0.330</td>
<td>0.550</td>
<td>0.588</td>
<td>0.256</td>
</tr>
<tr>
<td>β</td>
<td>1</td>
<td>1.482</td>
<td>0.172</td>
<td>0.512</td>
<td>1.475</td>
<td>0.136</td>
</tr>
<tr>
<td>8: T=200, λ=1, censoring=30%, positive (α+βz)=78%, ρxz=0.71, ρxu=0.23, ε~N(0,3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCGMM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>α</td>
<td>1</td>
<td>1.081</td>
<td>0.374</td>
<td>0.382</td>
<td>1.143</td>
<td>0.290</td>
</tr>
<tr>
<td>β</td>
<td>1</td>
<td>0.971</td>
<td>0.261</td>
<td>0.263</td>
<td>0.929</td>
<td>0.203</td>
</tr>
<tr>
<td>SCLS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>α</td>
<td>1</td>
<td>0.531</td>
<td>0.490</td>
<td>0.679</td>
<td>0.567</td>
<td>0.386</td>
</tr>
<tr>
<td>β</td>
<td>1</td>
<td>1.493</td>
<td>0.243</td>
<td>0.549</td>
<td>1.498</td>
<td>0.196</td>
</tr>
</tbody>
</table>
Appendix

Proof of Lemma 3.1

Based on the definition, \( f_t(\beta) = 1(z_t'\beta > 0) \cdot \min\{ \max\{ y_t - x_t'\beta, -z_t'\beta\}, z_t'\beta\} \cdot w_t \)

for any \( \beta, \beta^0 \) in \( \Theta \), the following inequalities hold,

\[
|f_t(\beta) - f_t(\beta^0)| \leq |x_t'(\beta - \beta^0)| \cdot |w_t| \\
\leq \|x_t\| \cdot \|w_t\| \cdot \|\beta - \beta^0\|,
\]

(3-1-1)

where \( f_t(\beta) \) is the i-th element of \( f_t(\beta) \) and \( \| \| \) is Euclidean norm. Following Definition 3.3, define \( L_t^0 = \|x_t\| \cdot \|w_t\| \).

Since 

\[
\lim_{n \to \infty} \sup \frac{1}{n} \sum_{t=1}^{n} E(L_t^0) = \lim_{n \to \infty} \sup \frac{1}{n} \sum_{t=1}^{n} E (\|x_t\| \cdot \|w_t\|) \\
\leq \lim_{n \to \infty} \sup \frac{1}{n} \sum_{t=1}^{n} E^{1/2} \|x_t\|^2 \cdot E^{1/2} \|w_t\|^2 \quad \text{(by Cauchy-Schwartz Inequality)} \\
\leq \lim_{n \to \infty} \sup \frac{1}{n} \sum_{t=1}^{n} E^{1/2} \|x_t\|^{2r} \cdot E^{1/2} \|w_t\|^{2r} \quad \text{(by Liapunov’s Inequality and } r > 2) \\
\leq \Delta^2 < \infty \quad \text{(by assumption DM)}
\]

Therefore, \{f_t(\beta)\} is almost surely Lipschitz-L_1 on \( \Theta \).

Q.E.D.

3.1a. Derivation of inequality (3-1-1):

Only two cases will be shown here, and other cases can be shown similarly.

First, when \( z_t'\beta > 0 \) and \( z_t'\beta^0 \leq 0 \), define \( \gamma = \beta - \beta^0 \), then \( z_t'\beta = z_t'\gamma + z_t'\beta^0 \). Since \( z_t'\beta^0 \leq 0 \) and \( z_t'\beta > 0 \), we get \( z_t'\beta + z_t'\beta^0 \leq z_t'\gamma + z_t'\beta^0 \), and then \( 0 < z_t'\beta \leq z_t'\gamma \), thus \( |z_t'\beta| \leq |z_t'\gamma| \).

\[
|f_t(\beta) - f_t(\beta^0)| = \min\{ \max\{ y_t - x_t'\beta, -z_t'\beta\}, z_t'\beta\} \cdot w_t, \text{ thus} \\
\]

i) if \( y_t - x_t'\beta \leq -z_t'\beta \), \( |f_t(\beta) - f_t(\beta^0)| = |z_t'\beta| \cdot |w_t| \).
ii) if \(-z_i^\beta<y_i-x_i^\beta<z_i^\beta\), then \(|y_i-x_i^\beta| < |z_i^\beta|\), and
\[ |f_i(\beta) - f_i(\beta^0)| = |y_i-x_i^\beta||w_i| < |z_i^\beta||w_i|.\]

iii) if \(z_i^\beta\leq y_i-x_i^\beta\), \(|f_i(\beta) - f_i(\beta^0)| = |y_i-x_i^\beta||w_i| < |z_i^\beta||w_i|.\)

Therefore, in any case, we have \( |f_i(\beta) - f_i(\beta^0)| < |z_i^\gamma||w_i|\leq |x_i^\gamma||w_i|.\)

Second, when \(z_i^\beta>0\) and \(z_i^\beta^0>0\),
\[ |f_i(\beta) - f_i(\beta^0)| = \min\{\max\{y_i-x_i^\beta, -z_i^\beta\}, z_i^\beta\} - \min\{\max\{y_i-x_i^\beta^0, -z_i^\beta^0\}, z_i^\beta^0\}||w_i|.\]

i) if \(-z_i^\beta< y_i-x_i^\beta< z_i^\beta\) and \(y_i-x_i^\beta^0< -z_i^\beta^0\), then
\[ |f_i(\beta) - f_i(\beta^0)| = |y_i-x_i^\beta + z_i^\beta^0||w_i|.\]

Based on condition i), we get
\[ -z_i^\beta + z_i^\beta^0< y_i-x_i^\beta + z_i^\beta^0< z_i^\beta + z_i^\beta^0 \quad \text{and} \quad y_i-x_i^\beta + z_i^\beta^0< x_i^\beta - x_i^\beta^0. \]

Thus, \(-z_i^\beta + z_i^\beta^0< y_i-x_i^\beta + z_i^\beta^0< x_i^\beta - x_i^\beta^0\). Hence we get either
\[ |f_i(\beta) - f_i(\beta^0)| \leq |x_i^\gamma||w_i|, \quad \text{or} \quad |f_i(\beta) - f_i(\beta^0)| \leq |z_i^\gamma||w_i| < |x_i^\gamma||w_i|.\]

ii) if \(y_i-x_i^\beta< -z_i^\beta\) and \(y_i-x_i^\beta^0< z_i^\beta^0\), then \(|z_i^\beta + z_i^\beta^0|<|x_i^\beta - x_i^\beta^0|\), thus
\[ |f_i(\beta) - f_i(\beta^0)| = |z_i^\beta + z_i^\beta^0||w_i| = |z_i^\beta + z_i^\beta^0||w_i| < |x_i^\gamma||w_i|.\]

This shows the inequality (3.1-1). Q.E.D.

**Proof of Lemma 3.2**

The proof that the elements of \(f_i(\beta)\) are near epoch dependent (NED) on \(\{w, x, v\}\) of size \(-1/2\) uniformly on \((\Theta, \rho)\) follows Theorem 4.5 and Theorem 4.8 (Gallant and White 1988).

Let the \(L_p\) norm of a random variable \(Z\) be denoted \(\|Z\|_p = E^{1/p}|Z|^p, p \geq 1\), and the conditional expectation \(E_{t-m}^{t+m}(\cdot) = E(\cdot | F_{t-m}^{t+m})\), where \(F_{t-m}^{t+m} = \sigma(x_{t-m} v_{t-m}, ..., x_{t+m} v_{t+m}).\)

3.2a **Show that \(y_i^*\) is NED on \(\{w, x, v\}\) of size \(-(r-1)/(r-2)\) for some \(r > 2\).**
\[ \left\| y_t^* - E_{t-m}^{t+m}(y_t^*) \right\|_2 = \left\| x_t^* + \beta_0^* - E_{t-m}^{t+m}(x_t^* + \beta_0^*) \right\|_2 \]
\[ = \left\| x_t^* + \beta_0^* - x_t^* + \beta_0^* - E_{t-m}^{t+m}(u_t^*) \right\|_2 \]
\[ = \left\| u_t^* - E_{t-m}^{t+m}(u_t^*) \right\|_2 \]
\[ = \nu_m. \text{ (by Assumption NE)} \]

Given Assumption NE, \( \{u_t^*\} \) are near epoch dependent on \( \{x_t, v_t\} \) (hence trivially on \( \{w_t, x_t, v_t\} \)) with \( \nu_m \) of size \(-\frac{(r-1)}{(r-2)}\) for \( r > 2 \), then \( y_t^* \) is NED on \( \{w_t, x_t, v_t\} \) of size \(-\frac{(r-1)}{(r-2)}\) for some \( r > 2 \).

### 3.2b Verify the domination conditions

For \( \beta' \) in \( \Theta \) and a constant \( \delta^0 \), there exists \( \eta^0(\delta^0) = \{\beta \in \Theta : \rho(\beta, \beta^0) \leq \delta^0\} \). Define

\[ \hat{y}_{mt}^* = E_{t-m}^{t+m}(y_t^*), \text{ then } \hat{y}_{mt}^* \text{ is measurable-F}_{t-m}, \text{ and } x_t \text{ is always measurable-F}_{t-m}, \text{ given that near epoch dependence is on } \{w_t, x_t, v_t\}, \hat{x}_{mt} = E_{t-m}^{t+m}(x_t) = x_t. \]

For every \( \beta \) in \( \eta^0(\delta^0) \), we have

\[ \left| f_t(y_t^*, x_t, w_t, \beta) - f_t(\hat{y}_{mt}^*, \hat{x}_{mt}, w_t, \beta) \right| \]
\[ = \left| 1(z_t^* > 0) \cdot w_t \cdot \left( (\min \{\max \{y_t^* - x_t^*; -z_t^*\beta\}, z_t^*\beta\}) - (\min \{\max \{\hat{y}_{mt}^* - x_t^*; -z_t^*\beta\}, z_t^*\beta\}) \right) \right| \]
\[ \leq \left| w_t \right| \cdot ( \left| y_t^* - \hat{y}_{mt}^* \right| + \left| x_t - \hat{x}_{mt} \right| + \left| w_t - \hat{w}_{mt} \right| ) \quad (3.2b-1) \]
\[ = B_t^0(w_t, \beta) \cdot d_t, \]

where \( B_t^0(w_t, \beta) = \left| w_t \right| \leq \left\| w_t \right\|, \text{ note that } \left\| \right\| \text{ denote Euclidean norm, and} \]
\[ d_t = \left| y_t^* - \hat{y}_{mt}^* \right| + \left| x_t - \hat{x}_{mt} \right| + \left| w_t - \hat{w}_{mt} \right| \]
\[ = \left| y_t^* - \hat{y}_{mt}^* \right|. \]

Given assumption DM, \( \left\| w_t \right\| \) is \( 2r \)-integrable, and \( B_t^0(w_t, \beta) \) is \( q \)-dominated, where \( q = 2r \) and \( r > 2 \), on \( \Theta \) uniformly. Next we need to show that \( B_t^0(w_t, \beta) \cdot d_t \) is \( r \)-dominated.
\[ \left\| B_t^\eta(w_t, \beta) \cdot d_t \right\|_p \leq \left\| B_t^\eta(w_t, \beta) \right\|_{L^r} \cdot \left\| d_t \right\|_{L^r} \]  
(by Cauchy-Schwartz inequality)

\[ \leq \left\| w_t \right\|_{L^r} \cdot \left\| y_t^* - y_{m^*} \right\|_{L^r} \]

\[ \leq \left\| w_t \right\|_{L^r} \cdot \left( \left\| y_t^* \right\|_{L^r} + \left\| y_{m^*} \right\|_{L^r} \right) \]  
(by Minkowski's inequality)

\[ \leq \left\| w_t \right\|_{L^r} \cdot \left( \left\| y_t^* \right\|_{L^r} + \left\| y_t^* \right\|_{L^r} \right) \]  
(Conditional Jensen's inequality) (3-2b-2)

\[ \leq 2\Delta \left\| x_t^* \beta_0 + u_t^* \right\|_{L^r} \]

\[ \leq 2\Delta \left( \left\| x_t^* \beta_0 \right\|_{L^r} + \left\| u_t^* \right\|_{L^r} \right) \]  
(by Minkowski's inequality)

\[ \leq 2\Delta \left( \left\| \beta_0 \right\| \cdot \left\| x_t^* \right\|_{L^r} + \left\| u_t^* \right\|_{L^r} \right) \]

\[ \leq 2(K+1)\Delta^2 \leq \infty \]  
(by compactness of \( \Theta \))

Thus, \( B_t^\eta(w_t, \beta) \cdot d_t \) is \( r \)-dominated for \( r > 2 \). The above inequality holds for any \( \beta^0 \) in \( \Theta \) and that \( \delta^0 \) may be arbitrary for \( \eta^\delta(\delta^0) \equiv \{ \beta \in \Theta : \rho(\beta, \beta^0) \leq \delta^0 \} \). So the function is independent of \( \beta^0 \), and \( \eta^\delta(\delta^0) \) can be the entire parameter space \( \Theta \). Hence \( B_t^\eta(w_t, \beta) \cdot d_t \) is \( r \)-dominated on \( \Theta \) uniformly, for \( t = 1, 2, \ldots \).

3.2c The size of NED

Based on Theorem 4.5 (Gallant and White 1988), \( \eta_{\text{mp}} = \sup_t \left\| d_t \right\|_p \) for \( p = 2r/(2r-1) \) and \( r > 2 \) is of size \(-2a(r-1)/(r-2)\). Since

\[ \eta_{\text{mp}} = \sup_t \left\| y_t^* - y_{m^*} \right\|_p \]

\[ = \sup_t \left\| u_t^* - E_{t-m}^{t+m}(u_t^*) \right\|_p \]

\[ \leq \sup_t \left\| u_t^* - E_{t-m}^{t+m}(u_t^*) \right\|_2 \]  
(3-2c-1)

\[ = \nu_m = O(m^\lambda) \]  
for all \( \lambda < -(r-1)/(r-2) \) and \( r > 2 \).
The inequality (3-2c-1) follows from Liapunov’s inequality and the fact that
\[ p = \frac{2r}{(2r-1)} < \frac{4}{3} < 2 \] for \( r > 2 \). Then \( \eta_{\text{top}} \) is of size \(-(r-1)/(r-2)\) for \( r > 2 \), and \( a = 1/2 \). Thus, the elements of \( f_i(\beta) \) are near epoch dependent (NED) on \( \{w_i, x_i, v_i\} \) of size \(-1/2\) on \( (\Theta, \rho) \).

3.2d Uniform NED

Since the function \( B_i^0 \) is independent of \( \beta^0 \) and \( \delta^0 \) can be chosen arbitrarily, thus the required inequalities hold over the entire parameter space. Therefore, based on Theorem 4.8, the elements of \( f_i(\beta) \) are indeed near epoch dependent (NED) on \( \{w_i, x_i, v_i\} \) uniformly on \( (\Theta, \rho) \).

The proof of Lemma 3.2 is complete. Detailed proofs for some inequalities are shown below.

3.2.1 Show the inequality (3-2b-1)

If \( z_t^\prime \beta < 0 \), the inequality holds trivially. For \( z_t^\prime \beta > 0 \), the following inequality is always true:

\[
\Psi = \left| (\min\{\max\{y_t^* - x_t^\prime \beta, -z_t^\prime \beta\}, z_t^\prime \beta\}) - (\min\{\max\{\hat{y}_{\text{mt}}^* - x_t^\prime \beta, -z_t^\prime \beta\}, z_t^\prime \beta\}) \right|
\leq \left| y_t^* - \hat{y}_{\text{mt}}^* \right|,
\]

where, \( \hat{y}_{\text{mt}}^* \) is the predicted values of \( y_t^* \) based on the “recent epoch”.

Only three cases will be shown here, and other cases can be shown similarly.

First, if \(-z_t^\prime \beta < y_t^* - x_t^\prime \beta < z_t^\prime \beta \) and \( \hat{y}_{\text{mt}}^* - x_t^\prime \beta < -z_t^\prime \beta \), then \( \Psi = \left| y_t^* - x_t^\prime \beta + z_t^\prime \beta \right| \). Since \( 0 < y_t^* - x_t^\prime \beta + z_t^\prime \beta \), and \( -\hat{y}_{\text{mt}}^* \) - x_t^\prime \beta \geq -z_t^\prime \beta \), thus \( \Psi \leq \left| y_t^* - \hat{y}_{\text{mt}}^* \right| \).

Second, if \(-z_t^\prime \beta < y_t^* - x_t^\prime \beta < z_t^\prime \beta \) and \( \hat{y}_{\text{mt}}^* - x_t^\prime \beta > z_t^\prime \beta \), then
\[
\Psi = \left| y_t^* - x_t^\prime \beta - z_t^\prime \beta \right| = \left| -y_t^* + x_t^\prime \beta + z_t^\prime \beta \right| \). Since \(-y_t^* + x_t^\prime \beta + z_t^\prime \beta > 0 \), and \( \hat{y}_{\text{mt}}^* > z_t^\prime \beta + x_t^\prime \beta \), then \( \hat{y}_{\text{mt}}^* - y_t^* > z_t^\prime \beta + x_t^\prime \beta - y_t^* \), so \( \Psi \leq \left| y_t^* - \hat{y}_{\text{mt}}^* \right| \).

Third, if \( y_t^* - x_t^\prime \beta < -z_t^\prime \beta \) and \( \hat{y}_{\text{mt}}^* - x_t^\prime \beta > z_t^\prime \beta \), then \( \Psi = \left| 2z_t^\prime \beta \right| \). Since \(-y_t^* + x_t^\prime \beta > z_t^\prime \beta \), thus
\[ y_{mt}^* - y_i^* > 2z'_i \beta > 0, \text{ so } \Psi = \left| 2z'_i \beta \right| \leq \left| y_i^* - y_{mt}^* \right|. \]

Therefore, the inequality (3-2b-1) follows immediately because \( |x_i - \hat{x}_{mt}| = 0. \)

3.2-2 \textit{Show the inequality (3-2b-2)}

By conditional Jensen’s inequality, we have
\[
\left| E_{t-m} \left( y_i^* \right) \right|^{2r} \leq E \left[ (y_i^*)^{2r} \mid F_{t-m} \right],
\]
\[
E^{1/(2r)} \left| E_{t-m} \left( y_i^* \right) \right|^{2r} \leq E^{1/(2r)} \left[ (y_i^*)^{2r} \mid F_{t-m} \right]
\]
\[
= E^{1/(2r)} (y_i^*)^{2r} \quad \text{(by law of iterated expectation).}
\]

Q. E. D.

\textbf{Proof of Lemma 3.3}

Let \( S_n^0 \) be an open sphere in \( \Theta_n \) centered at \( \beta_0 \) with fixed radius \( \varepsilon > 0 \). For each \( n \), define the neighborhood \( \zeta_n^0(\varepsilon) = S_n^0(\varepsilon) \cap \Theta_n \) with compact complement \( \zeta_n^0(\varepsilon)^c \) in \( \Theta_n \). Since the minimum of the limiting GMM function \( Q_n(\beta) \) is zero, based on the condition (2.8), the moment condition \( M(\beta) = 0 \) is satisfied at \( \beta_0 \), thus \( \beta_0 \) is the minimizer for \( Q_n(\beta) \) in \( \zeta_n^0(\varepsilon) \). Given Assumption ID iii), we have \[ \left\{ \frac{1}{n} \sum_{t=1}^{n} E(z'_t \mid y_t > 0) \right\} > 0 \] for every \( \beta \) in \( \zeta_n^0(\varepsilon) \). Hence for any other \( \beta \) in \( \zeta_n^0(\varepsilon) \) for all \( \varepsilon > 0 \), \( M(\beta) \neq 0 \) and \( \lim \inf_{n \to \infty} \left\{ \min_{\beta \neq \beta_0} Q_n(\beta) - Q_n(\beta_0) \right\} > 0 \), and thus \( \beta_0 \) is the unique minimizer for \( Q_n(\beta) \) in \( \zeta_n^0(\varepsilon) \). Since the radius \( \varepsilon \) of \( S_n^0 \) can be chosen to be large enough such that the complement \( \zeta_n^0(\varepsilon)^c \) is empty, therefore, \( \beta_0 \) is the unique minimizer for \( Q_n(\beta) \) in \( \Theta_n \).

Q. E. D.

\textbf{Proof of Theorem 3.4}

The existence is immediate following Gallant and White (1988) Theorem 2.2. The proof of consistency follows Theorem 3.3 and Theorem 3.19 (Gallant and White, 1988). The continuity
of $Q_n(\beta)$ on $\Theta$ follows from the continuity of $f_i(\beta)$, which is straightforward. The identifiably unique minimizer $\beta_o$ for $Q_n(\beta)$ on $\{\overline{\Theta}_n\}$ is based on Lemma 3.3. In addition, the random constraint sequence $\Theta_n$ will convergence to $\overline{\Theta}_n$.

Next, the convergence of $Q_n(\beta)$ to $Q_0(\beta)$ in probability uniformly for all $\beta$ follows the ULLN of Gallant and White (1988, Theorem 3.18). Given Lemma 3.1 and Lemma 3.2, we only need to show that the elements of $f_i(\beta)$ are r-dominated on $\Theta$ uniformly in $t=1,2,\ldots, r \geq 2$, which, among others, allows us to define $M(\beta) = \frac{1}{n} \sum_{i=1}^{n} E [f_i(\beta)]$, by ensuring the expectations exist.

$$\left| f_n(\beta) \right| < \left\| f_i(\beta) \right\|$$

$$\leq [z't\beta \cdot w_i'w_i]^{1/2} \quad (3-3-1)$$

$$\leq [(x't\beta) \cdot w_i'w_i]^{1/2}$$

$$\leq \left\| x_i \right\| \left\| w_i \right\| \left\| \beta \right\|$$

$$\leq K \left\| x_i \right\| \left\| w_i \right\|, \quad \text{(by compactness of } \Theta)$$

for some $K > 0$. Define $D_t = K \left\| x_i \right\| \left\| w_i \right\|$, then

$$E^{1/r} D_t^r = KE^{1/r} \left( \left\| x_i \right\| \left\| w_i \right\| \right)^r$$

$$\leq KE^{1/(2r)} \left\| x_i \right\|^{2r} \cdot E^{1/(2r)} \left\| w_i \right\|^{2r} \quad \text{(by Cauchy-Schwartz Inequality)}$$

$$\leq K \Delta^2 < \infty. \quad \text{(by assumption DM)}$$

Thus the elements of $f_i(\beta)$ are r-dominated on $\Theta$ uniformly in $t = 1,2,\ldots, r \geq 2$.

Therefore, Theorem 3.4 holds following Theorem 3.19 (Gallant and White, 1988).

3.3a The inequality (3-3-1) can be shown as following:

$$\left\| f_i(\beta) \right\| = [(z_i'\beta)^2 \cdot w_i'w_i]^{1/2} \quad \text{if} \quad y_i-x_i'\beta \geq z_i'\beta \text{ or } y_i-x_i'\beta \leq -z_i'\beta$$

or

$$= [(y_i - x_i'\beta)^2 \cdot w_i'w_i]^{1/2} \quad \text{if} \quad -z_i'\beta < y_i - x_i'\beta < z_i'\beta.$$
Since \(-z_i'\beta < y_i - x_i'\beta < z_i'\beta\), then \(\left| y_i - x_i'\beta \right|^2 < \left| z_i'\beta \right|^2\). Thus inequality (3-4-1) is immediate.

Q.E.D.

**Proof of Lemma 4.1**

The proof follows the CLT for near epoch dependent function of a mixing process given by Theorem 5.3 (Gallant and White, 1988). As defined in the text,

\[ B_n^0 = \text{Var}(\sqrt{n} F_n(\beta_0)) \]

\[ = \text{Var}(n^{1/2} \sum_{i=1}^{n} f(\beta_0)). \]

4.1a. proving that \(f(\beta_0)\) is NED of size \(-1\), uniformly on \((\Theta, \rho)\) follows the same procedure in proving Lemma 3.2.

4.1b. Show that \((B_n^0)^{1/2}\) is \(O(1)\).

Given the assumption of uniform positive definiteness of \(B_n^0\), we only need to show \(B_n^0\) is \(O(1)\). Define \(Z_{nt} = \lambda' f(t(\beta_0)),\) then \(E(Z_{nt}) = E(\lambda' f(t(\beta_0))) = 0\), where \(\lambda \in R^l, \lambda'\lambda = 1\) for arbitrary \(\lambda\) so that \(\text{Var}(n^{1/2} \sum_{t=1}^{n} Z_{nt}) = \lambda' B_0^0 \lambda.\) Now

\[ \text{Var}(n^{1/2} \sum_{t=1}^{n} Z_{nt}) = n^{-1} E\left(\sum_{t=1}^{n} Z_{nt}\right)^2 \]

\[ \leq n^{-1} E\left[ \max_{1 \leq j \leq n} \left(\sum_{t=1}^{n} Z_{nt}\right)^2 \right] \quad (4-1b-1) \]

Since \(f(\beta_0)\) is NED of size \(-1\), \(Z_{nt} = \lambda' f(\beta)\) is NED of size \(-1\) (a fortiori of size \(-1/2\)).

We have shown that the elements of \(f(\beta)\) are \(r\)-dominated on \(\Theta\), then \(\left\| Z_{nt} \right\|_r \leq \Delta < \infty\). Thus, following lemma 3.14 (Gallant and White 1988), \(\{Z_{nt}\}\) is a mixingale of size \(-1\), with

\[ C_{nt} = \text{Max} (\left\| Z_{nt} \right\|_r, 1) \leq \Delta < \infty \text{ for all } n \text{ and } t. \]
Applying McLeish’s inequality (Theorem 3.11, Gallant and White 1988), and from inequality 4-1b-1, we have

\[ \text{Var} \left( n^{-1/2} \sum_{t=1}^{n} Z_{nt} \right) \leq n^{-1} K \left( \sum_{t=1}^{n} C_{t} \right) \]

\[ \leq n^{-1} K(n\Delta^2) \]

\[ = K\Delta^2 < \infty, \text{ where } K \text{ is a finite constant.} \] Then

\[ \text{Var} \left( n^{-1/2} \sum_{t=1}^{n} Z_{nt} \right) = \lambda' B_{n}^0 \lambda \leq K\Delta^2 < \infty. \]

Thus, \{\lambda' B_{n}^0 \lambda\} is O(1) for arbitrary \( \lambda \in \mathbb{R}, \lambda' \lambda = 1 \), implying that \{B_{n}^0\} is O(1), and \((B_{n}^0)^{-1/2}\) is also O(1).

4.1c Apply the Central Limit Theorem

Now define \( Z_{nt} = \lambda' (B_{n}^0)^{-1/2} f_t(\beta_0) \), where \( \lambda \in \mathbb{R}, \lambda' \lambda = 1 \). Since \( f_t(\beta_0) \) is r-dominated and \((B_{n}^0)^{-1/2}\) is O(1), then \( Z_{nt} \) is r-dominated. \( \|Z_{nt}\| \leq \Delta < \infty, \text{ for } r > 2. \)

\[ V_n^2 = \text{Var} \left( \sum_{t=1}^{n} Z_{nt} \right) \]

\[ = \text{Var} \left( \sum_{t=1}^{n} (\lambda' (B_{n}^0)^{-1/2} f_t(\beta_0)) \right) \]

\[ = \lambda' (B_{n}^0)^{-1/2} \text{Var} \left( \sum_{t=1}^{n} f_t(\beta_0) (B_{n}^0)^{-1/2} \lambda \right) \]

\[ = \lambda' (B_{n}^0)^{-1/2} (\lambda' (B_{n}^0)^{-1/2} \lambda) \]

\[ = n. \]

Thus, \( V_n^2 = n^{-1} \), which is O(n^{-1}) as the CLT required.

Since \( E(Z_{nt}) = \lambda' (B_{n}^0)^{-1/2} E(f_t(\beta_0)) = 0 \) and \( f_t(\beta_0) \) is NED of size \(-1\) (and then \{Z_{nt}\} is NED of size \(-1\)), based on the CLT, we get

\[ V_n^{-1} \sum_{t=1}^{n} Z_{nt} \xrightarrow{d} N(0,1). \]
4.1d Limiting distribution of $F_n(\beta_0)$

We have shown that $n^{-1/2} \sum_{i=1}^{n} (\lambda_i' (B_n^0)^{-1/2} f_i(\beta_0) \xrightarrow{d} N(0,1)$. For arbitrary $\lambda \in \mathbb{R}^l, \lambda' \lambda = 1$,

by the Cramér–Wold device (Theorem 25.6, Davidson, 1994), we get

$$n^{-1/2} \sum_{i=1}^{n} (B_n^0)^{-1/2} f_i(\beta_0) \xrightarrow{d} N(0, I).$$

Thus,

$$(B_n^0)^{-1/2} \sqrt{n} F_n(\beta_0) \xrightarrow{d} N(0, I).$$

Q.E.D.

**Proof of Lemma 4.2**

The proof for Lemma 4.2 and Theorem 4.3 follows Theorem 7.2 and 7.3 in Newey and McFadden (1994). Although Theorem 7.2 and 7.3 are based on the i.i.d. assumption, they can be extended to NED case here. As can be seen from the proof of Theory 7.2 and 7.3, the dependence structure of the moment function does not need to be specified as long as the consistency condition and the condition (iv) in Theorem 7.2 (pp. 2186) are satisfied. The consistency of the SCGMM estimator has been shown in section III, and Lemma 4.1 above shows that the condition (iv) is satisfied for the NED moment function here. Hence, the stochastic equicontinuity condition can be used here to show the asymptotical distribution.

Based on the model, $f_i(\beta_0) = 1(z_i' \beta_0 > 0) \left[ u_i^* \cdot 1(-z_i' \beta_0 < u_i^* < z_i' \beta_0) + z_i' \beta_0 \cdot 1(u_i^* \geq z_i' \beta_0) - z_i' \beta_0 \cdot 1(u_i^* \leq -z_i' \beta_0) \right]$. Define $\gamma = \beta - \beta_0$, thus

$$f_i(\beta) = 1(z_i' \beta_0 + z_i' \gamma > 0) \left[ (u_i^* - x_i' \gamma) \cdot 1(-z_i' \beta_0 + x_i' \gamma - z_i' \gamma < u_i^* < z_i' \beta_0 + x_i' \gamma + z_i' \gamma) + z_i' \beta \cdot 1(u_i^* \geq z_i' \beta_0 + x_i' \gamma + z_i' \gamma) - z_i' \beta \cdot 1(0 < u_i^* < z_i' \beta_0 + x_i' \gamma - z_i' \gamma) \right].$$

Define $\Delta f_i(w_i, x_i) = 1(z_i' \beta_0 > 0) \left\{ -x_i' \cdot 1(-z_i' \beta_0 < u_i^* < z_i' \beta_0) + z_i' \cdot [1(u_i^* \geq z_i' \beta_0) - 1(u_i^* \leq -z_i' \beta_0)] \right\}$. 

38
4.2a Show \( r(x, w, \beta) \to 0 \) with probability one.

\[
r(x, w, \beta) = \|f(\beta) - f(\beta_0) - \Delta f(w, x)\| (\beta - \beta_0) ||\beta - \beta_0||
\]

\[= \|1(z'_0 + z'_\gamma > 0) w_i \cdot (u_i^* - x_i') 1(z'_0 < u_i^* < z'_0 + x_i' + z'_\gamma)
    + (z'_0 + z'_\gamma) [1(0 < u_i^* < z'_0 + x_i' + z'_\gamma) - 1(z'_0 < u_i^* < z'_0 + x_i' + z'_\gamma)]
    + 1(z'_0 > 0) w_i \cdot (u_i^* + x_i') 1(-z'_0 < u_i^* < z'_0)
    - (z'_0 + z'_\gamma) [1(z'_0 + x_i' + z'_\gamma > 0) 1(u_i^* > 0, z'_0)]
\]

For all \( \beta \) in a suitably small neighborhood of \( \beta_0 \), if \( z'_0 > 0 \), then \( z'_0 > 0 \); and if \( z'_0 < 0 \), then \( z'_0 < 0 \).

In the latter case, \( r(x, w, \beta) = 0 \). If \( z'_0 > 0 \), and assume \( z'_\gamma > 0 \) (if \( x'_\gamma < 0 \), it can be shown similarly), and also by assumption \( x'_\gamma - z'_\gamma > 0 \), we get

\[
r(x, w, \beta)
\]

\[= ||1(z'_0 > 0) w_i \cdot (u_i^* - x_i') 1(z'_0 < u_i^* < z'_0 + x_i' + z'_\gamma)
    + (z'_0 + z'_\gamma) [1(0 < u_i^* < z'_0 + x_i' + z'_\gamma) - 1(z'_0 < u_i^* < z'_0 + x_i' + z'_\gamma)]
    + 1(z'_0 > 0) w_i \cdot (u_i^* + x_i') 1(-z'_0 < u_i^* < z'_0)
    - (z'_0 + z'_\gamma) [1(z'_0 + x_i' + z'_\gamma > 0) 1(u_i^* > 0, z'_0)]||\|\gamma\|
\]

\[= ||1(z'_0 > 0) w_i \cdot (u_i^* - x_i' - z'_0 - z'_\gamma) 1(z'_0 < u_i^* < z'_0 + x_i' + z'_\gamma)
    + (u_i^* + x_i') - z'_0 - z'_\gamma) 1(-z'_0 < u_i^* < z'_0 + x_i' + z'_\gamma) - 1(z'_0 < u_i^* < z'_0 + x_i' + z'_\gamma)||\|\gamma\|
\]

\[\leq ||1(z'_0 > 0) w_i \cdot (x_i' + z'_\gamma) 1(z'_0 < u_i^* < z'_0 + x_i' + z'_\gamma)
    + 1(-z'_0 < u_i^* < z'_0 + x_i' + z'_\gamma))||\|\gamma\|
\]

\[\leq ||1(z'_0 > 0) w_i \cdot (2x_i') 1(z'_0 < u_i^* < z'_0 + x_i' + z'_\gamma)
    + 1(-z'_0 < u_i^* < z'_0 + x_i' + z'_\gamma))||\|\gamma\|
\]

\[\leq 2\|w_i\|\|x_i\| \cdot (1(z'_0 < u_i^* < z'_0 + x_i' + z'_\gamma) + 1(z'_0 < u_i^* < z'_0 + x_i' + z'_\gamma))
\]

(by Holder’s inequality, Liapunov’s inequality)
Therefore, for any sufficient small $\delta_n$, $E[\sup_{\|w\|=\delta} r(x, w, \beta)] = E(2\|w_t\| \|x_t\|) < \infty$.

Moreover, as $\beta \to \beta_0$ (i.e. $\gamma \to 0$), $\text{Prob}(u_t^* = \pm z_t' \beta_0) = 0$, thus with probability one, $r(x, w, \beta) \to 0$.

4.2b Show that

$$\frac{1}{n} \sum_{i=1}^{n} \Delta f_i(w_t, x_t) \xrightarrow{p} \frac{1}{n} \sum_{i=1}^{n} E(\Delta f_i(w_t, x_t)).$$

For the $ij$-th element of $\Delta f_i(w_t, x_t)$,

$$|\Delta f_{ij}(w_t, x_t)| = |1(z_t' \beta_0 > 0) [- (w_t x_t')_j - 1(-z_t' \beta_0 < u_t^* < z_t' \beta_0)]$$

$$+(w_t z_t')_j 1(u_t^* \geq z_t' \beta_0) - 1(u_t^* \leq -z_t' \beta_0)|.$$

In any case, it is true that for $r > 2$, $\sup_{t} \|\Delta f_{ij}(w_t, x_t)\| \leq \Delta^2 < \infty$ (by Cauchy-Schwartz inequality and Assumption DM). Thus each element of $\Delta f_i(w_t, x_t)$ is $L_r$ bounded uniformly in $t$, for $r > 2$.

Thus, the convergence of $\frac{1}{n} \sum_{i=1}^{n} \Delta f_i(w_t, x_t)$ to $\frac{1}{n} \sum_{i=1}^{n} E(\Delta f_i(w_t, x_t))$ follows a Law of Large Numbers for near epoch dependent processes (Davidson, 1994, Theorem 20.19, page 324).

Therefore, following Theorem 7.3 (Newey and McFadden 1994), Lemma 4.2 holds.

Q.E.D.

**Proof of Theorem 4.3**

The proof follows Theorem 7.2 and Theorem 7.3 (Newey and McFadden, 1994). Related material can also be found in Andrews (1994a). Given the limiting distribution and the stochastic equicontinuity condition in Lemma 4.1 and Lemma 4.2, we only need to show that

$$G_n^0 = \frac{1}{n} \sum_{i=1}^{n} E[\Delta f_i(w_t, x_t)],$$

where $G_n^0$ is defined in the text and $\Delta f(w_t, x_t)$ is defined in proof of Lemma 4.2.

$$\frac{1}{n} \sum_{i=1}^{n} E[\Delta f_i(w_t, x_t)]$$
\[ \frac{1}{n} \sum_{t=1}^{n} E_{x,w} \left[ \Delta f_t (w_t, x_t) \mid (x,w) \right] \]

\[ = \frac{1}{n} \sum_{t=1}^{n} E_{x,w} \left[ w_t \right. \{ -x_t' \cdot 1(-z_t' \beta_0 < u_t^* < z_t' \beta_0) \]

\[ + z_t'^* \left[ 1(u_t^* \geq z_t' \beta_0) - 1(u_t^* \leq -z_t' \beta_0) \right] \mid (x,w) \} \]

\[ = \frac{1}{n} \sum_{t=1}^{n} E_{x,w} \left[ w_t \cdot \left( -z_t'^* \int_{-\infty}^{-z_t' \beta_0} g(\lambda) d\lambda \right. \right. \]

\[ - x_t' \int_{-z_t' \beta_0}^{z_t' \beta_0} g(\lambda) d\lambda + z_t'^* \int_{z_t' \beta_0}^{\infty} g(\lambda) d\lambda \mid (x,w) \] \]

\[ = \frac{1}{n} \sum_{t=1}^{n} E \left[ 1(- z_t' \beta_0 < u_t^* < z_t' \beta_0) (-w_t, x_t') \right], \]

which is \( G_n^0 \).

Therefore, following Theorem 7.2 (Newey and McFadden 1994), Theorem 4.3 holds.

Q.E.D.

**Proof of Theorem 5.1**

We need to show \( \hat{G}_n (\hat{\beta}_n) \xrightarrow{p} G_n^0 \) and \( \hat{B}_n (p_n) \xrightarrow{p} B_n^0 \) separately.

5.1a The consistence of \( B_n^0 \)

Following the same procedures before, we can show that the elements of \( \{ f(\beta) \} \) are NED on \( \{ x_i, v_i \} \) of size \(-2(r-1)/(r-2)\), and they are 2r-dominated on \( \Theta \) uniformly in \( t = 1, 2, ..., \) and \( r > 2 \).

Thus, following Lemma 6.6, Lemma 6.7, and Theorem 6.8 (Gallant and White 1988),

\( \hat{B}_n (p_n) \xrightarrow{p} B_n^0 \) is immediate.

5.1b The consistency of \( \hat{G}_n (\hat{\beta}_n) \)
Since \( \hat{G}_n(\beta_k) = \frac{1}{n} \sum_{t=1}^{n} 1(-z_t' \hat{\beta}_k < u_t < z_t' \hat{\beta}_k) (-w_t x_t') \), define

\[
G_n(\beta_0) = \frac{1}{n} \sum_{t=1}^{n} 1(-z_t'\beta_0 < u_t < z_t'\beta_0) (-w_t x_t').
\]

We first show that each element of the difference \( \hat{G}_n(\beta_k) - G_n(\beta_0) \) converges to zero in probability, given the consistency of \( \hat{\beta}_k \). Then we will show that \( G_n(\beta_0) \) converges to \( E(G_n(\beta_0)) \). Define \( \gamma_n = \hat{\beta}_k - \beta_0 \), for each element in \( G_n(\hat{\beta}_k) \),

\[
\left| [G_n(\hat{\beta}_k) - G_n(\beta_0)]_{ij} \right| = \left| 1(-z_i' \hat{\beta}_k < u_i < z_i' \hat{\beta}_k)(-w_i x_i')_{ij} - 1(-z_i'\beta_0 < u_i^* < z_i'\beta_0)(-w_i x_i')_{ij} \right|
\]

\[
= \left| 1(-z_i'\beta_0 + x_i'\gamma_n - z_i'\gamma_n < u_i^* < z_i'\beta_0 + x_i'\gamma_n + z_i'\gamma_n)(-w_i x_i')_{ij}
- 1(-z_i'\beta_0 < u_i^* < z_i'\beta_0)(-w_i x_i')_{ij} \right|
\]

\[
\leq \|w_i\| \|x_i\| \cdot \left| 1(-z_i'\beta_0 + x_i'\gamma_n - z_i'\gamma_n < u_i^* < z_i'\beta_0 + x_i'\gamma_n + z_i'\gamma_n) - 1(-z_i'\beta_0 < u_i^* < z_i'\beta_0) \right|
\]

Because \( \hat{\beta}_k \to \beta_0 \) by the consistency of \( \hat{\beta}_k \), then \( \gamma_n \to 0 \). Given the domination condition for \( w_i \) and \( x_i \), these terms converge to zero in probability. Thus, \( \hat{G}_n(\beta_k) \to P G_n(\beta_0) \). The convergence of \( G_n(\beta_0) \) converges to \( E(G_n(\beta_0)) \) follows a Law of Large Numbers for near epoch dependent processes (Davidson, 1994, Theorem 20.19, page 324). Therefore, \( \hat{G}_n(\beta_k) \) is a consistent estimator of \( G_n^0 \).

Q.E.D.

**Proof of Corollary 5.2**

It is immediate following Hansen (1982), Davidson and MacKinnon (1993, Chapter 17).